Structure-function studies of key amyloid-degrading enzymes

Moss, Stephen

Award date:
2020

Awarding institution:
University of Bath

Structure-function studies of key amyloid-degrading enzymes

Stephen Moss

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

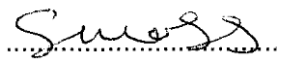
Department of Biology and Biochemistry

August 2020

recognise that its copyright rests with the author and that they must not copy it or use material from it except as permitted by law or with the consent of the author.

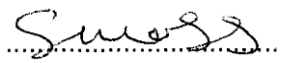
Declaration of any previous submission of the work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.



Declaration of authorship

I am the author of this thesis, and the work described therein was carried out by myself personally.



Acknowledgements

Firstly, I would like to thank my supervisors Prof. Ravi Acharya and Dr Vasanta Subramanian for their guidance and support over the course of this Ph.D. project. It has been a joy to work with people with great knowledge and enthusiasm for their respective fields.

I would like to thank the Alzheimer's Society for their funding, without which, this research would not have been possible. It has been a pleasure to meet members and volunteers of the organization. They have certainly brought a greater context to my research and provided me a huge amount of motivation.

The most exciting times I have enjoyed during my research have been at Diamond Light Source. I am grateful for the opportunity to have access to the facility and to all the beamline scientists who have helped me over the past three years.

I would like to thank all the members of the Acharya group whom I have been fortunate enough to meet over the last three years. Special thanks must go to Gyles Cozier from whom I have learnt plenty. Jonathon Davies and William Bradshaw were also greatly helpful in getting to grips with the field of X-ray crystallography.

I am enormously grateful for the support my Mother, Father, Sister and Grandparents who have provided me support over so many years. Finally, I would like to thank my partner Meagan - you have made the last three years particularly enjoyable. I appreciate all your help especially in proofreading this thesis - I am sure I will get to repay the favour soon enough.

Abstract

The accumulation of the small peptide amyloid-β has been proposed as a major trigger for the development of Alzheimer's disease. Within the brain, the concentration of amyloid-β is tightly controlled through production and clearance mechanisms. Studies have revealed that reduced levels of amyloid-β clearance are present in individuals living with Alzheimer's disease which results in the observed accumulation of the peptide. This accumulation of amyloid-β can lead to the formation of large aggregated amyloid plaques - one of two detectable hallmarks of the disease.

Amyloid degrading enzymes (ADEs) are a group of proteins with the ability to degrade amyloid-β and are major players in the clearance of amyloid-β. Stimulating ADE activity or expression, in order to compensate for the decreased clearance in the Alzheimer's disease phenotype, provides a promising therapeutic target. It has been reported in mice that upregulation of ADEs can reduce the levels of amyloid-β peptide and amyloid plaques – in some cases this led to improved cognitive function.

Among several known ADEs, neprilysin (NEP) and endothelin-converting enzyme I (ECE-1) have been identified to be two of the most potent. Both proteins have the capacity to digest soluble amyloid-β which, in turn, can form the toxic oligomeric species found within the brains of individuals living with Alzheimer's disease. Whilst NEP and ECE-1 are known for their amyloid-degradation, they exhibit complexity through the additional substrates that they can degrade. Two such groups of substrates are the endothelins and natriuretic peptides. Both peptide groups are vasoactive and illustrate the strong link between NEP, ECE-1 and blood pressure regulation. The promiscuous nature of NEP and ECE-1 raises questions surrounding their specificity and how such a broad range of substrates can be degraded.

This research builds upon current structural and functional understanding of both NEP and ECE-1, giving a detailed insight into the molecular interactions that leads to the hydrolysis of peptide substrates. This research has used a cross-disciplinary approach of X-ray structure

determination, mutagenesis and kinetic analysis in order to elucidate the crucial tasks performed by key ADEs.

Contents

List of figures

List of tables

Abbreviations

A_{280}	Absorbance at 280nm
ACE	Angiotensin-converting enzyme
ADE	Amyloid-degrading enzyme
ANP	Atrial natriuretic peptide
APP	Amyloid precursor protein
BNP	B-type natriuretic peptide
CNP	C-type natriuretic peptide
Cryo-EM	Cryo electron microscopy
CHICUP	Cross-linking of unmodified proteins
DLS	Diamond Light Source
D_{max}	Maximum particle dimension
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
E. coli	*Escherichia coli*
ECE-1	Endothelin-converting enzyme I
GSH	Glutathione
GSSG	Glutathione disulfide
HEK 298T	Human embryonic kidney 298T cells
IDE	Insulin-degrading enzyme

LB	Lysogeny broth
MAD	Multiwavelength anomalous dispersion
NEP	Neprilysin
NMR	Nuclear magnetic resonance
P. pastoris	*Pichia pastoris*
PBS	Phosphate buffered saline solution
PDB	Protein data bank
PICUP	Photo-Induced cross-linking of unmodified proteins
R_g	Radius of gyration
SAXS	Small-angle X-ray scattering
SD	Standard deviation
XRC	X-ray crystallography

Standard abbreviations for amino acids, nucleotides, SI units/prefixes, and chemical elements are used throughout.

Chapter 1 – Introduction

1.1 Dementia, Alzheimer's disease and amyloid-β

1.1.1 Dementia

Dementia is a syndrome which encompasses a broad range of brain diseases. Whilst the first pathologies of dementia were reported in 1906, much of the past century has seen symptoms of dementia confused with the normal process of aging. However, an increasingly aging population combined with effective outreach work conducted by several UK and international charities have helped to bring dementia into the spotlight. Dementia has emerged as a prominent threat to the worlds aging populations and, while recent focus has drawn both funding and attention to the syndrome, much remains to be understood about the basics of this disease.

According to the world health organisation "Dementia is a syndrome in which there is deterioration in memory, thinking, behaviour and the ability to perform everyday activities". Improved awareness and diagnosis have begun reveal the impact of dementia within the UK, where dementia is now recognised as the leading cause of death (ONS, 2015). Current estimates predict over 1 million people will be living with dementia in the UK by 2025 (Prince *et al.,* 2015). In addition to the social impact dementia has on individuals and families, the syndrome presents a substantial financial burden. In 2014 the cost of dementia in the UK was estimated at £26 billion with a large part of this money spent on the many years of social care required for individuals living with advanced forms of dementia.

Clarity on the global impact of dementia is also becoming evident with estimates that 46.8 million people worldwide were living with dementia in 2015. These figures are expected to rise greatly particularly in low to middle income countries as life expectancy increases. These lower income countries will be less well equipped to deal with the large amount of care required for individuals with dementia (Prince *et al.,* 2015). The far-reaching implications of dementia means there can be no doubt that it represents one of the largest hurdles for science in the 21st century.

1.1.2 Alzheimer's disease

By far the most prevalent form of dementia is Alzheimer's disease, which accounts for some 60-70% of global cases. Alzheimer's disease can be separated into familial or sporadic forms. Familial forms of the disease account for around 2-3% of total cases and are the result of

genetic predispositions to the disease (Campion *et al.,* 1999). Typically, familial Alzheimer's disease is early onset with the first symptoms appearing in individuals in their 30's and 40's. The sporadic form of Alzheimer's disease accounts for all remaining cases of the disease and the first symptoms of the disease do not usually appear until over 65 years of age (Bird, 1999).

The symptoms of Alzheimer's disease was first described by Alois Alzheimer in 1906 following his investigation into the disease progression of a patient displaying memory loss, disturbed sleep, aggressiveness and confusion (Stelzma *et al.,* 1995). Alzheimer recorded the presence of a distinct brain pathology to the disease, namely the presence of neurofibrillary tangles and senile plaques. These tangles and plaques remain identifiable hallmarks to the disease over 100 years later (Maurer *et al.,* 1997).

More recent research into the Alzheimer's disease pathology has revealed the composition of both hallmarks. Senile plaques are composed of extracellular deposits of the small peptide fragment amyloid-β (Glenner and Wong, 1984), while neurofibrillary tangles are composed of hyperphosphorylated tau protein (Wood *et al.,* 1986). These two hallmarks have shaped the field of Alzheimer's disease research and represent core components in proposed hypotheses describing disease progression (Kumar *et al.,* 2015).

The most recognised Alzheimer's disease hypothesis, the amyloid cascade hypothesis, was published in 1992 (Hardy *et al.,* 1992). It proposed that the principle component in development of Alzheimer's disease was amyloid-β peptide. An imbalance in production or degradation of amyloid-β led to its extracellular aggregation. This accumulation then leads to neuronal cell death directly or indirectly through downstream effects including neurofibrillary tangles. While the hypothesis has changed over the past two decades the core principle remains the same and the evidence linking amyloid-β to Alzheimer's disease remains strong (Selkoe and Hardy, 2016).

1.1.3 Evidence for the role of amyloid-β in Alzheimer's disease

Much of the research into Alzheimer's disease has been conducted through investigations into individuals with familial Alzheimer's disease (FAD). The inherited genetic mutations that cause FAD provide a means of investigating the cause of the disease. Missense mutations in PSEN1 and PSEN2 are one of the prevalent causes of FAD (Bird, 1999) with these genes

encoding the protein γ-secretase which is responsible for amyloid-β processing. Mutations in the PSEN genes can lead to elevated levels of higher molecular weight amyloid-β species which in turn have a higher propensity for aggregation (Hardy, 1997) - amyloid-β lengths will be discussed in greater detail below. Other FAD causing mutations have also been identified in or directly flanking the amyloid precursor protein (APP) gene which encodes amyloid-β (Goate *et al.,* 1991; Citron *et al.,* 1992). Conversely, missense mutations in the APP gene can result in reduced amyloid-β expression and provide a protective function against Alzheimer's disease (Rovelet-Lecrux *et al.,* 2006; Sleegers *et al.,* 2006).

Additional evidence for the role of amyloid-β in Alzheimer's disease has been provided through research into individuals with trisomy of chromosome 21 (Down's syndrome). These individuals possess a third copy of the APP gene and display symptoms similar to those observed in FAD at very early ages (Oyama *et al.,* 1994; Head *et al.,* 2012). Indeed, in rare cases where individuals without Down's syndrome possess a duplication of just the APP gene, similar FAD symptoms are observed (Prasher *et al.,* 1998).

A large focus has been placed on studying the genetic factors of Alzheimer's disease development in individuals with FAD or Down's syndrome to understand the underlying mechanism of Alzheimer's disease. While this is a good basis for research, the vast majority of Alzheimer's disease cases are sporadic in nature, occurring in individuals above the age of 65. Analysis of individuals with sporadic Alzheimer's disease has revealed one of the highest predictors of disease development is the possession of the APOE4 gene. Indeed, over 40% of individuals with Alzheimer's disease possess this gene. The protein transcribed from APOE4, ApoE4, disrupts amyloid-β clearance across the blood-brain barrier to a greater extent than alternative isoforms ApoE3 and ApoE2 (Deane *et al.,* 2008).

Whilst the evidence for the role of amyloid-β in Alzheimer's disease is strong, there is several pieces of conflicting evidence. Perhaps the most significant of these is the lack of Alzheimer's disease associated cognitive decline in many elderly people that possess large quantities of amyloid-β plaques (Davis *et al.,* 1999; Fagan *et al.,* 2009). A commonly used counter argument to this evidence is that large amyloid-β plaques appear to be non-toxic, and it is the smaller amyloid-β oligomers that can cause neuronal cell death.

Another frequently used case against the role of amyloid-β in Alzheimer's disease is the disappointing results from clinical trials aiming to reduce amyloid-β load. γ-secretase inhibitors have been unsuccessful in combatting the cognitive decline of Alzheimer's disease (Doody *et al.,* 2013; De Strooper, 2014). In fact, increased cognitive decline was observed in phase III trials in addition to serious side effects including skin cancer (Henley *et al.,* 2014). More recent therapeutic efforts targeting Alzheimer's disease have been focused on reducing or clearing the amyloid-β load in the brain. Several monoclonal antibodies including Bapineuzumab and Solanezumab have been used to target monomeric, oligomeric or aggregated forms of amyloid-β. While these antibodies were able to reduce the amyloid-β load, they failed to prevent cognitive decline in phase III clinical trials (van Dyck, 2018).

The majority of Alzheimer's disease therapeutics have failed to reduce cognitive decline in clinical trials for reasons that can be addressed. Secretase inhibitors were bad drug targets due to the additional substrates γ-secretase is known to cleave. Their inhibition therefore resulted in unforeseen and undesirable side effects. Likewise, while results for antibody therapies have been disappointing, a consensus in opinion suggests that drugs were used too late in the development of the disease. Molecular changes in the brain are suggested to have occurred decades before cognitive impairment was observed (Selkoe and Hardy, 2016).

Despite the number unsuccessful trials for Alzheimer's disease therapeutics there has been a promising recent development for an antibody-based therapy. While phase III drug trials for monoclonal antibody Aducanumab were prematurely ended due to lack of significant results, a recent post hoc analysis has revealed that at high dosage there may be a small but significant reduction to cognitive decline. While the effects of this antibody are small, it may provide a strong proof of principle for the continued use of amyloid-β as a key therapeutic target.

1.1.4 Amyloid-β production and processing

Amyloid-β is formed through proteolytic processing of amyloid precursor protein (APP). APP, a type 1 transmembrane glycoprotein, is expressed ubiquitously across the central nervous system and peripheral tissues (Slunt *et al.,* 1994). Alternative splicing of the APP

mRNA generates eight isoforms with sizes ranging from 695 to 770 amino acids. Of these the APP_{695} isoform is the most prevalent within the central nervous system. Translation of APP occurs in the endoplasmic reticulum (ER) and from there the protein exhibits complex post translational processing and trafficking (Kang *et al.,* 1987; Dyrks *et al.,* 1988; Weidemann *et al.,* 1989). N-linked and O-linked glycosylation, phosphorylation and tyrosine sulfation modifications are made within the Golgi apparatus. It is here that much of protein remains, with just a small fraction being trafficked further to the plasma membrane. The fraction of APP present on the cell surface can be cleaved by several enzymes to release soluble protein fragments or can be rapidly reinternalized and trafficked into lysosomes for degradation (Haass *et al.,* 1992). The proteolytic processing and trafficking of APP has been comprehensively reviewed by Haass *et al*. (2012) and Jiang *et al*. (2014).

APP contains three cleavage sites that are targeted by transmembrane proteins α-, β- and γ-secretase. Initial cleavage can be made by either α- or β-secretase - α-secretase cleavage commits APP to non-amyloidogenic processing, while cleavage by β-secretase leads to amyloidogenic processing and the formation of amyloid-β. β-Secretase displays optimal activity in an acidic environment and has been suggested to be rate limiting for the formation of amyloid-β (Knops *et al.,* 1995). Studies indicate that β-secretase is trafficked from the ER to the plasma membrane through the Golgi apparatus and endosomes. It is here that most APP cleavage has been suggested to take place due to the lower pH conditions within the acidic lumen (Vassar, 1999; Vassar, 2001). The overlapping trafficking of APP and β-secretase allows amyloid-β production to occur in the ER, Golgi and endosomes in addition to the cell surface (Cook *et al.,* 1997; Hartmann *et al.,* 1997).

Cleavage of APP by β-secretase results in the release of N-terminal soluble APPβ (sAPPβ) fragment while the C-terminal 99 amino acid fragment remains within the membrane (Figure 1.1a). The C-terminal fragment can then be cleaved by γ-secretase to release amyloid-β (Weidemann *et al.,* 2002). γ-Secretase forms an amyloid-β fragment 48 or 49 amino acids depending on the site of cleavage. The amyloid-β fragment is then cleaved in a stepwise manner by γ-secretase (Figure 1.1b). Due to the nature of this cleavage process amyloid-β is produced with lengths between 37-49. Of these, the 40- and 42-residue species are most prevalent (Bolduc *et al.,* 2016; Takahashi *et al.,* 2002).

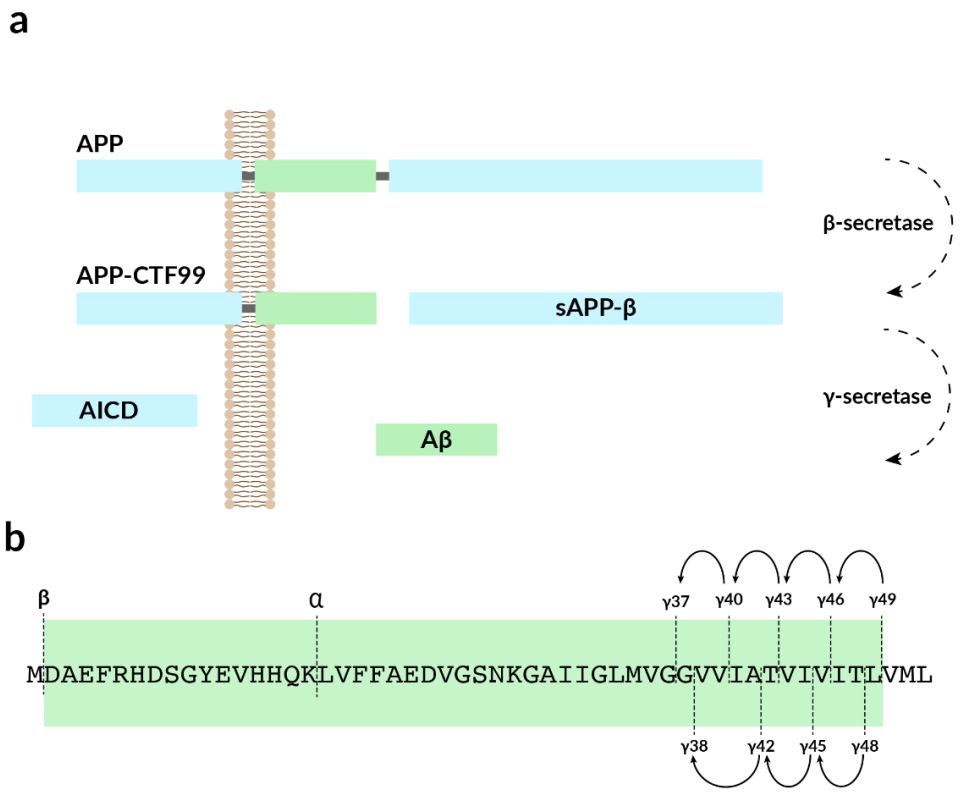


Figure 1.1, Amyloid-β protein processing. (a) Amyloid precursor protein (APP) is processed by β- and γ- secretase enzymes to release amyloid-β into the extracellular space. (b) The amyloid-β sequence (green) is shown with cleavage sites for α-, β- and γ- secretase enzymes. γ-secretase cleavage occurs in a sequential manner, removing two, three or four amino acid residues at a time.

The biological roles of APP and amyloid-β have yet to be fully characterised but some functions have been suggested. APP has been proposed to be involved in cell-cell adhesion through dimerization (Soba *et al.,* 2005) and cell-matrix adhesion through interaction with heparin (Clarris *et al.,* 1997), collagen (Beher *et al.,* 1996) and laminin (Kibbey *et al.,* 1993). Additional roles have been suggested in neuronal development (Young-Pearse *et al.,* 2007), signalling (Tamayev *et al.,* 2009) and intracellular transport (Goldstein, 2012). Van Der Kant and Goldstein (2015) provide a summary of the current research relating to the function of APP.

Amyloid-β is suggested to act as regulatory peptide, modulating synaptic activity and neuronal survival (Pearson and Peers, 2006). It has also been suggested that amyloid-β may act as a transcription factor (Bailey *et al.,* 2011). Due to the evidence for a physiological role of both APP and amyloid-β, complete elimination of either protein is considered an undesirable outcome for Alzheimer's disease therapies.

1.1.5 Structures of amyloid-β

The proteolytic processing of APP results in a range of amyloid-β peptides that vary in length. Once formed these peptides can exist in monomeric, oligomeric or fibrillar forms. Structural determination of each form is critical for understanding the molecular role of amyloid-β in Alzheimer's disease. A detailed review of the amyloid-β structures determined by nuclear magnetic resonance (NMR) and X-ray crystallographic (XRC) is provided by Chen *et al*. (2017), while (Viola and Klein, 2015) review the links between amyloid-β species and Alzheimer's disease pathogenesis. Below these topics are covered more briefly.

Amyloid-β monomers are soluble peptides that have been largely reported to be helical in nature. An NMR structure of amyloid-β 1-42 (1IYT) (Crescenzi *et al.,* 2002) revealed the presence of two α-helical fragments between residues 8-25 and 27-38 (Figure 1.2). These helices are connected by a β-turn and are flanked by disordered N- and C-termini. Likewise, an NMR structure of amyloid-β 1-40 (1BA4) (Coles *et al.,* 1998) reports the peptide to be helical in nature but with slight variation in the residues contributing to the helices (15-25 and 27-36). Structures of smaller amyloid-β fragments have also been reported to adopt similar conformations - an NMR structure of amyloid-β 1-28 adopts a helical conformation (Talafous *et al.,* 1994).

It is important to note that the NMR structures discussed above were not determined in biologically relevant solvents due to the propensity of amyloid-β to aggregate at the concentrations required for structural investigation. Because of this, other techniques have been used to support NMR data. Molecular dynamic studies propose that amyloid-β adopts a majority α-helical or β-sheet conformations with rapid transitions between the two (Yang and Teplow, 2008). The Ile41 and Ala42 residues present in amyloid-β 1-42 are suggested to shift the equilibrium towards a β-sheet conformation, causing an increase in propensity for aggregation when compared with amyloid-β 1-40.

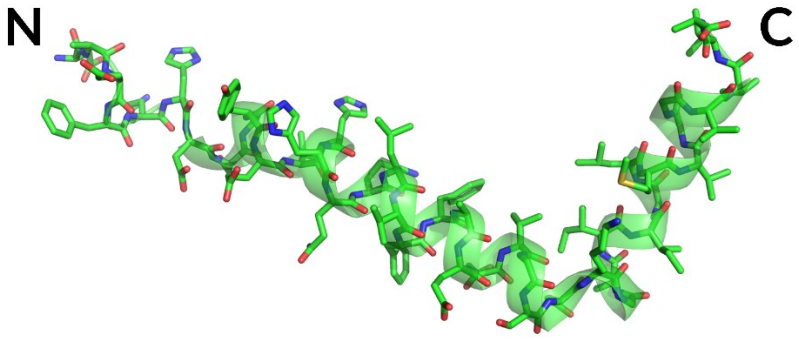


Figure 1.2, NMR structure of monomeric amyloid-β 1-42. The structure of amyloid-β (1IYT) adopts an α-helical conformation with a β-turn within the centre of the peptide. Molecular dynamics research has suggested that amyloid-β exhibits a rapid transition between α-helical or β-sheet conformations

Amyloid-β oligomers are a highly diverse range of species generated from variation in the length, number and spatial arrangement of amyloid-β monomers. Like monomers, oligomers are soluble, however, due to their unstable and heterogenous nature they are inherently difficult to study. Crosslinking of amyloid-β peptides within oligomers has provided a means of covalently stabilising the oligomeric species. Methods such as PICUP (Bitan *et al.,* 2003) and CHIPUP (Williams *et al.,* 2016) allow crosslinking of peptides without substrate modification. Crosslinking has revealed that oligomeric species in solution can adopt dimers, trimers, tetramers, pentamers, hexamers and heptamers with varying degree. It was also revealed that amyloid-β 1-42 has a higher propensity for forming high mass oligomers when compared with amyloid-β 1-40. While crosslinking studies have provided some insight into these oligomers, it is unclear as to the impact that crosslinking has on the stability of species formed.

Structural studies have yet to yield conclusive high-resolution information of oligomer structures. Despite limited information there is growing evidence to suggest that cyclic oligomeric species may play the most crucial role in Alzheimer's disease. Lambert *et al*. (1998) first noticed cyclic oligomeric species with the use of electron microscopy. More recent NMR data supports the formation of cyclic pentamers and hexamers that are able to stack to form decamers and dodecamers (Ahmed *et al*. 2010).

While the previous research on amyloid-β oligomers is substantial, full characterisation remains incomplete. Attempting to determine atomic resolution structures represents a

major challenge, not least because of the practical difficultly in dealing with the oligomers, but also in finding a system that accurately represents the natural environment. Further complexity is also present when considering that oligomers will be formed in both intra- and extracellular environments (Jung Lee *et al.,* 2017).

Amyloid fibrils are insoluble unbranched fibres formed through the stacking of amyloid-β monomers. This species of amyloid-β can assemble into large extracellular plaques – one of the major pathogenic hallmarks for Alzheimer's disease.

Amyloid-β fibrils possess in register β-sheets that run down the long axis of the fibril offering high stability. This stability has made structural studies easier when compared with the oligomeric form. To date, several structures of fibrils have been solved using NMR and cryo-EM and their polymorphic nature has become clear (Figure 1.3). NMR structures have revealed fibrils of amyloid-β 1-42 can form through stacking of both monomeric (Xiao *et al.,* 2015) and dimeric (Colvin *et al*. 2016) amyloid-β units. A trimeric form has also been reported with amyloid-β 1-40 (Lu *et al.,* 2013).

Additional variation has been reported in the residues that are buried or surface exposed within a fibril (Schmidt *et al.,* 2015; Colvin *et al.,* 2016; Wälti *et al.,* 2016; Gremer *et al.,* 2017). A large contributor to this is the size of the amyloid-β monomer. Amyloid-β 1-42 contains Ala42 which is able to form a salt bridge with Lys28 offering an explanation for the higher propensity of amyloid-β 1-42 to aggregate compared with amyloid-β 1-40 (Xiao *et al.,* 2015).

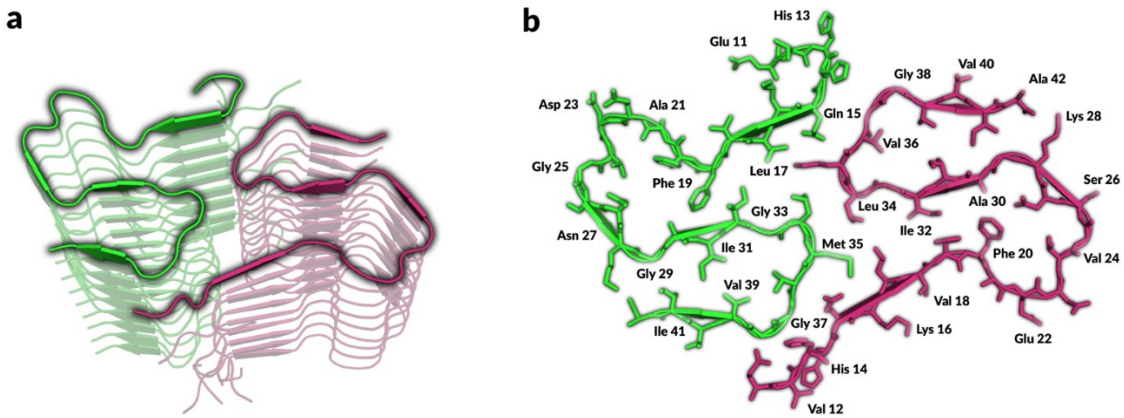


Figure 1.3, NMR structure of amyloid-β fibrils. (a) NMR structure of monomorphic amyloid-β fibril 5KK3. Single amyloid-β monomers stack to form parallel β-sheets down the long axial plane. (b) Two amyloid-β monomers

form a dimer through side chain interactions and together form the stacking unit for fibril formation. Peptide disorder at the N-terminus (residues 1-12) are not modelled due to their highly disordered nature.

1.1.6 Amyloid-β – peptide toxicity

When the amyloid cascade hypothesis was published in 1992 it proposed that the primary cause of Alzheimer's disease was the large aggregated plaques formed from amyloid-β fibrils (Hardy and Higgins, 1992). Studies in the years following have revealed the poor correlation between amyloid-β plaques and neuronal toxicity. Indeed, neurotoxicity of amyloid-β was reported to be present without the formation of fibrils at all (Oda *et al.,* 1994; Hsia *et al.,* 1999; Mucke *et al.,* 2000). More recently amyloid-β oligomers have been proposed as the primary mechanism by which neuronal toxicity was caused. The last two decades has seen this hypothesis develop and there is now considerable evidence implicating oligomers in neuronal toxicity both *in vitro* and *in vivo* (Walsh *et al.,* 2002a).

Oligomers appear to accumulate in both intracellular (Baker-Nigh *et al.,* 2015) (Zheng *et al.,* 2013), and extracellular (Takeda *et al.,* 2013) pools with movement between localisations. Intracellular oligomers show no preferred localisation to specific organelles (Zheng *et al.,* 2013), while extracellular oligomers localise to the neuronal surfaces (Gong *et al.,* 2003) in what may be a cell type specific interaction (Lambert *et al.,* 2001; Lacor *et al.,* 2007). This extracellular localisation suggested the presence of a cell type specific membrane protein that, through interaction with amyloid-β oligomers, may result in neurnal toxicity. Many membrane proteins are possible targets and have been reviewed by Viola and Klein, (2015) and Jung Lee *et al.* (2017). It may be that through membrane protein interaction amyloid-β oligomers downstream effects are mediated such as tau hyperphosphorylation (De Felice *et al.,* 2007).

1.2 Amyloid-degrading enzymes

Control of amyloid-β levels is achieved through production and clearance. In Alzheimer's disease this process becomes imbalanced. It has been suggested that individuals with early-onset familial Alzheimer's have an increase in amyloid-β production (Hatami *et al.,* 2017). In contrast, the more common sporadic form of Alzheimer's disease has been proposed to be

a result of reduction in amyloid-β clearance (Mawuenyega *et al.,* 2010). Understanding of these clearance mechanisms are therefore critical understanding of Alzheimer's disease.

Amyloid-β clearance mechanisms can broadly be separated into two groups - enzymatic and non-enzymatic. Non-enzymatic mechanisms include interstitial fluid drainage (Weller *et al.,* 2000), uptake by microglial phagocytosis (Frautschy *et al.,* 1998; Wyss-Coray *et al.,* 2003; Nielsen *et al.,* 2009) and transport across the blood brain barrier (Shibata *et al.,* 2000; Deane *et al.,* 2003). The formation of insoluble plaques may also be considered a clearance method by acting as a sink removing more toxic oligomeric amyloid-β species (Cheng *et al.,* 2007; Lesné *et al.,* 2008).

Enzymatic clearance of amyloid-β is achieved through the protease activity of amyloid-degrading enzymes (ADEs). These proteases function to digest amyloid-β peptides into smaller less-toxic species. While the majority of degradation is suggested to occur within the brain, clearance may also occur in other areas in which amyloid-β is localised after clearance from the brain. A diverse range of ADE have been discovered from several protease classes including metallo, serine, aspartyl, cysteine and threonine. Many of these proteases display activity towards multiple peptide substrates across a range of tissue localisations. The localisations and activities of ADE's have been reviewed by Saido *et al.* (2012) and Nalivaeva *et al.* (2012).

Amyloid-degrading metalloproteases	Reference
Neprilysin	(Howell *et al.,* 1995a)
Neprilysin 2	(Marr and Hafez, 2014)
Human membrane metallo-endopeptidase-like protein	(Huang *et al.,* 2008)
Endothelin-converting enzyme I	(Eckman *et al.,* 2001)
Endothelin-converting enzyme 2	(Eckman *et al.,* 2003)
Angiotensin-converting enzyme	(Hemming and Selkoe, 2005)
Insulin-degrading enzyme	(Qiu *et al.,* 1997)
Matrix metalloproteinase 2	(Roher *et al.,* 1994)
Matrix metalloproteinase 9	(Backstrom *et al.,* 1996)
CD147	(Vetrivel *et al.,* 2008)

Table 1.1, Amyloid-degrading enzymes from the metalloprotease family. All proteins have demonstrated an ability to degrade amyloid-β in vitro, but the biological relevance of each protein has yet to be fully characterised.

The best studied ADEs are members of the metalloproteases class (Table 1.1). Metalloproteases are the most diverse class of protease which can be separated into two groups based on cleavage location - exopeptidase (EC 3.4.17) and endopeptidase (EC 3.4.24). Enzymes from this class share a conserved mechanism based around a divalent cation. Much of the research into the mechanism of peptide cleavage by metalloproteases was conducted on thermolysin (Pelmenschikov *et al.,* 2002). The proposed water dependant mechanism for peptide cleavage is shown in Figure 1.4.

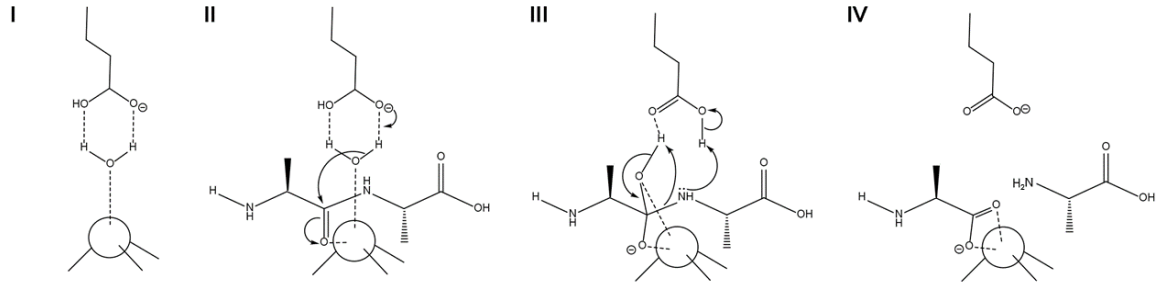


Figure 1.4, Metalloprotease catalytic mechanism. Four steps are involved in the proposed mechanism for water dependant peptide scissile bond cleavage. The divalent metal cation is displayed as a circle with three lines representing the coordination sphere. (I) a water molecule coordinates the divalent cation and is stabilised through hydrogen bonds to a catalytic glutamate residue. (II) substrate binding occurs, and additional divalent coordination is provided by the oxygen of the substrate. The polarised water molecule then attacks the scissile carbonyl group resulting in the formation of a tetrahedral intermediate. (III) Scissile bond breakage and double proton transfer occurs. (IV) The resulting peptide fragments can then be released (Cerdà-Costa and Gomis-Rüth, 2014).

Of the proteases listed in Table 1.1, four zinc metalloproteases – neprilysin, endothelin-converting enzyme I, insulin-degrading enzyme and angiotensin-converting enzyme - have received the majority of attention for their actions as ADEs (Figure 1.5). Crucially each enzyme has *in vivo* evidence supporting their role in amyloid-β degradation. Each of the four proteins displays commonality in a number of structural features, and each has a remarkable number of substrates that vary in sequence, size, shape and charge of substrate. Below each protein will be briefly described with a summary of evidence supporting their role as ADEs.

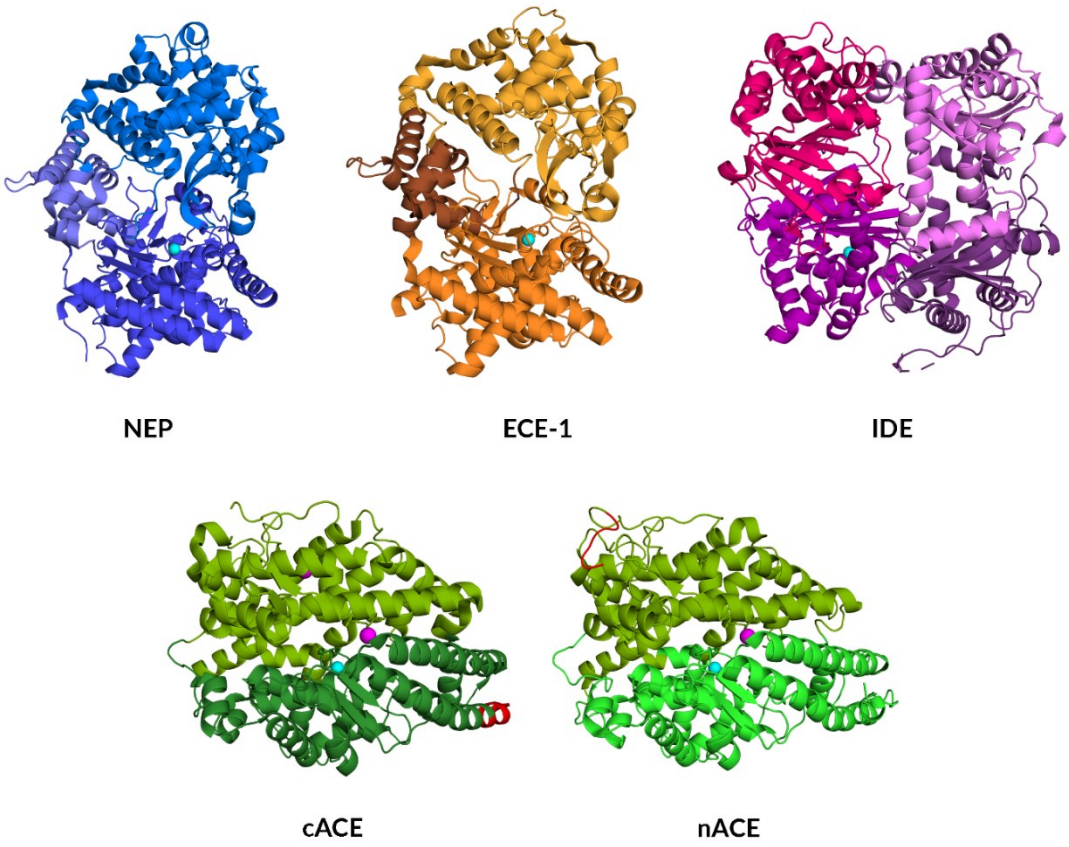


Figure 1.5, Crystal structures of principal ADEs. Neprilysin, endothelin-converting enzyme, insulin-degrading enzyme and N- and C- domains of angiotensin converting enzyme have been identified as important enzymes in amyloid-β regulation.

1.2.1 Neprilysin

Neprilysin (NEP, EC 3.4.24.11) is a transmembrane M13 zinc metalloprotease and the best known ADE (Iwata *et al.,* 2001). While NEP is expressed ubiquitously, high levels have been identified in the kidney and intestines (Uhlén *et al.,* 2015). After translation NEP is glycosylated before being transported to the plasma membrane where the large catalytic domain is presented into the extracellular space (Nalivaeva and Turner, 2013). Here the protein is known to degrade many small peptide substrates (<5 kDa) including substance P, angiotensin II, angiotensin I, bradykinin, enkephalins (Gafford *et al.,* 1983; Mumford *et al.,* 1981), neurotensin, oxytocin (Almenoff *et al.,* 1981; Johnson *et al.,* 1984) and amyloid-β. Initial HPLC analysis of NEP confirmed the protein could degrade amyloid-β *in vitro* and at several independent cleavage sites (Howell *et al.,* 1995b). Further mass spectrometry

analysis has provided greater resolution to the exact nature of amyloid-β cleavage (Rogeberg *et al.,* 2014). Presently it is clear that NEP can degrade monomeric amyloid-β however there remains an absence of evidence for oligomer or fibril degradation (Leissring, 2008).

In addition to the *in vitro* evidence, *in vivo* evidence for the key role of NEP in amyloid-β degradation has been building over the past two decades. A study from 2002 revealed NEP found within the brain is localised along axons and at synapses. Varied levels of expression were seen in different neuronal cell types and lower neprilysin expression correlated with higher amyloid-β deposition. The study also suggested that ageing causes a gradual down regulation of ADE and therefore led to the increase in amyloid-β seen in Alzheimer's disease (Fukami *et al.,* 2002).

Knockout mice models have also been used to demonstrate the importance of NEP in amyloid-β regulation. NEP knockout mice display up to a two-fold increase in amyloid-β 1-40 and 1-42 levels when compared with the wild-type in a range of brain tissues (Iwata *et al.,* 2001; Farris *et al.,* 2007). Specifically, levels of oligomeric amyloid-β have been shown to rise and impair cognitive function (Huang *et al.,* 2006). Transgenic mouse studies using mice overexpressing the APP gene have also supported NEP's role in amyloid-β regulation. These transgenic mice suffer abnormally high premature death rates (Hsiao *et al.,* 1995; Mucke *et al.,* 2000), however by crossing these mice with mice overexpressing NEP, there was a significant reduction in amyloid-β levels and the associated cytopathology (Leissring *et al.,* 2003).

While the majority of studies strongly implicate NEP in amyloid-β degradation and indicate a protective role for the enzyme, it should be noted there is contrasting evidence. Experiments using APP transgenic mice revealed NEP overexpression reduced the soluble monomeric amyloid-β levels but oligomeric levels were unchanged with no reduction in the cytopathology associated with increase amyloid-β levels (Meilandt *et al.,* 2009). The contrasting evidence may illustrate the complexity of the system due to the range of species that amyloid-β can adopt.

1.2.2 Endothelin-converting enzyme I

Endothelin-converting enzyme I (ECE-1, EC 3.4.24.71), like NEP, is a transmembrane zinc metalloprotease and member of the M13 metalloprotease family. In humans, ECE-1 shares 38% amino acid sequence similarity with NEP but exists as a disulfide linked dimer instead of as a monomer (Schulz *et al.,* 2009).

ECE-1 is expressed in four different isoforms aECE-1, bECE-1, cECE-1 and dECE-1. All isoforms are produced by the same gene, but variation in promoter leads to an altered N-terminal tail sequence. The four isoforms have differing subcellular localisations, but all retain the same level of substrate catalysis and efficiency. aECE-1 is localised to the plasma membrane, dECE-1 and cECE-1 localised to the plasma membrane and intracellular compartments and bECE-1 is localised almost entirely intracellularly (Valdenaire *et al.,* 1995; Schweizer *et al.,* 1997; Valdenaire *et al.,* 1999). These isoforms and their localisations have been reviewed by Eckman *et al*. (2001) and Pacheco-Quinto *et al*. (2013).

ECE-1 was initially characterised for its role in converting pro-endothelin/big-ET into endothelins – potent vasoconstrictors. Like NEP, ECE-1 is also expressed ubiquitously across all vascular cells and a range of tissues nonvascular cells including the lungs, kidneys, adrenal gland, ovary and testis (Xu *et al.,* 1994; Davenport *et al.,* 1998; Korth *et al.,* 1999). Across this range of tissues many substrates for ECE-1 have been identified in addition to pro-endothelin. These include bradykinin, neurotensin, substance P and amyloid-β (Johnson *et al.,* 1999; Eckman *et al.,* 2001). Optimal catalytic activity of ECE-1 varies for different substrates. For big ET-1 optimal activity is achieved at neutral pH while a pH of ~5.6 is optimal for bradykinin, substance P and amyloid-β. The localisation and pH optima for ECE-1 suggests that the majority of amyloid-β degradation by this enzyme is performed in the intracellular space (Saido and Leissring, 2012). Like NEP, this amyloid-β degradation has been identified to be specific to the monomeric species (Leissring, 2008).

The role of ECE-1 in Alzheimer's disease, while less well studied than NEP, has been confirmed through both *in vitro* and *in vivo* degradation of amyloid-β. *In vitro* studies have revealed ECE-1 expressing cell lines have a 2-3-fold increase in extracellular amyloid-β when treated with metalloprotease inhibitor phosphoramidon (Eckman *et al.,* 2001), this affect was proposed to stem from inhibition of intracellular degradation. Similarly, overexpressing

ECE-1 in CHO cells lacking endogenous ECE-1 expression reduced extracellular amyloid-β by 90% compared to a control. This amyloid-β reduction could be eliminated with phosphoramidon treatment. The *in vitro* work is supported by a transgenic mouse study that revealed increased brain amyloid-β levels in ECE-1 deficient mice (Eckman *et al.,* 2003).

1.2.3 Other notable ADEs

Presently, there are over 20 known ADEs (Nalivaeva *et al.,* 2014). While this is a large number of proteins, it is important to consider the physiological relevance of each one. A distinction should be made between enzymes that exhibit protease activity on amyloid-β in a test tube, where concentrations are greatly in excess of physiological conditions, and those that likely are responsible for amyloid-β degradation in a physiologically relevant system. *In vivo* evidence remains the best measure of the true relevance of an ADE. While NEP and ECE-1 are possibly the best-known ADEs, two other enzymes have substantial evidence for their role in Alzheimer's disease.

Insulin degrading enzyme (IDE, EC 3.4.24.56) is a 110kDa endopeptidase and a member of the M16 zinc metalloprotease family. IDE expression is ubiquitous but expression levels are tissue dependant (Bondy *et al.,* 1994). The post-translational processing and trafficking of the protein remain unclear but studies have revealed IDE is localised both intra- and extracellularly with a small fraction associating to the cell membrane (Raffaella Tundo *et al.,* 2012). IDE was initially characterised for its insulin degrading properties (Duckworth *et al.,* 1972), but has since been revealed to degrade a broad range of peptide targets including insulin, amylin, glucagon and amyloid-β (Shen *et al.,* 2006).

IDE has been identified as a major protease of amyloid-β in neuronal cell lines *in vitro* and *in vivo* mouse models. *In vitro* studies have revealed that IDE is able to digest intra- and extracellular amyloid-β pools (Qiu *et al.,* 1997; Qiu *et al.,* 1998; Sudoh *et al.,* 2002). Other studies present convincing evidence indicating that soluble monomeric amyloid-β is readily digested, but that oligomeric species are not digested by IDE (Walsh *et al.,* 2002b).

In vivo studies on mouse IDE double knockouts reveal cerebral accumulations of amyloid-β supporting IDE's role in amyloid-β digestion. Membrane fractions of brain cells and intact primary neuronal cells from these knockout mice were determined to have a >50% decrease in amyloid-β degradation using labelled amyloid-β (Vekrellis *et al.,* 2000; Farris *et al.,* 2003).

Transgenic mice overexpressing IDE show the reverse, with significant reduction in amyloid-β levels (Leissring *et al.,* 2003).

Angiotensin-converting enzyme (ACE, EC 3.4.15.1) is a zinc-dependant dipeptidyl carboxypeptidase best known for its role in blood pressure regulation through cleavage of angiotensin I and bradykinin (Ignjacev-Lazich *et al.,* 2005). Other substrates include enkephalins, chemotactic peptide, neurotensin and substance P (Skidgel and Erdös, 1987). Two isoforms of ACE are known – somatic ACE (sACE) and testicular (tACE) (Howard *et al.,* 1990). Both isoforms are transcribed from the same gene but while sACE is ubiquitous tACE is only expressed in mature testis (Igic and Behnia, 2003). The two ACE isoforms possess identical intracellular, transmembrane and catalytic domains (cACE) (Figure 1.6). The variation in isoforms stems from the additional N-terminal ACE domain (nACE) present in sACE but not in tACE. This additional catalytic domain has a high degree of structural similarity to cACE but substrate specificity is known to differ for some substrates (Natesh *et al.,* 2003; Corradi *et al.,* 2006).

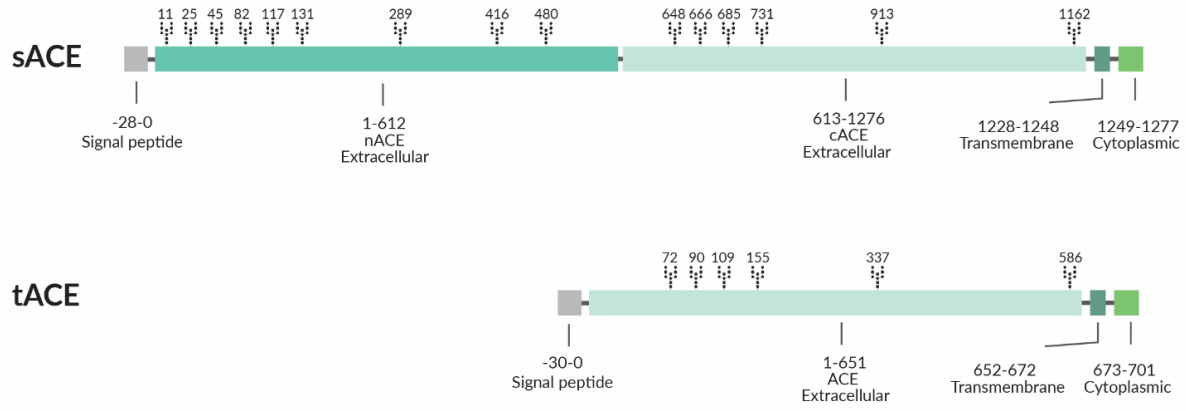


Figure 1.6, Domain organisation of somatic and testicular ACE. Both sACE and tACE possess identical cytoplasmic, transmembrane and cACE domains. The additional nACE domain is present in sACE has a high degree of sequence and structural homology to cACE.

Evidence for the role of ACE in Alzheimer's disease was revealed in the association between mutations and deletions within the ACE encoding DCP1 gene and an individual's susceptibility to Alzheimer's disease (Hu *et al.,* 1999; Kehoe *et al.,* 1999). Indeed, more recent

research has also flagged up the gene encoding ACE as an Alzheimer's disease susceptibility gene (Bertram *et al.,* 2007).

In vitro evidence for sACE as a ADE has been shown by the proteins ability to significantly inhibit aggregation and fibril formation of synthetic amyloid-β. This capacity can be removed by treatment with ACE specific inhibitor lisinopril (Hu *et al.,* 2001). While *in vivo* studies for ACE are inconclusive (Eckman *et al.,* 2006), the strong genetic link between ACE and Alzheimer's disease means the enzyme is regarded as an important ADE, but one that requires greater clarification.

1.2.4 Possible therapeutic targets of ADE benefit

It is clear that research into ADEs is critical to gain an understanding of the landscape of Alzheimer's disease as a whole. While many questions surrounding ADEs have yet to be addressed, current understanding has caused them to be identified as possible therapeutic targets for Alzheimer's disease treatment (Nalivaeva *et al.,* 2012; Graham *et al.,* 2017; Nalivaeva and Turner, 2019). If amyloid-β is a major driver of Alzheimer's disease, and if accumulation of amyloid-β is due to a reduction in clearance rather than an increase in production, a clear direction for therapy would be to compensate for the reduced clearance capacity. To this end, research into the upregulation of ADE expression (Saito *et al.,* 2005) or downregulation of ADE endogenous inhibitors (Jacobsen *et al.,* 2008) has been explored to increasing amyloid-β degradation. The potential to use gene therapy for increased ADE expression has been explored in mice (Marr *et al.,* 2003; Hong *et al.,* 2006). Such studies have demonstrated gene therapy to be effective means to reduce the amyloid pathology associated with AD. While results appear positive, the limitations for gene therapy remain a large hurdle with issues of secondary toxicity, exacerbation of the immune response, low efficiency and high cost (Li *et al.,* 2015).

Another obstacle that must be addressed if ADEs are to be considered viable therapeutic targets is their activity on diverse populations of substrates across a broad range of tissues. Upregulation of ADEs would no doubt generate unwanted secondary effects, the most likely of which would be the alteration in blood pressure regulation. Site directed mutagenesis is a possible means to mitigate issues of specificity. Through mutagenesis it has been possible

to change specificities of ADEs for substrates and tailor these proteins for amyloid-β degradation (Sexton *et al.,* 2012).

While the therapeutic potential of ADEs remains unclear, the necessity to understand amyloid-β clearance mechanisms has emerged as a critical area for Alzheimer's disease research as a whole. It may be through this that more effective treatment can be attained.

1.2.5 Other functions of ADEs

Given the large number of peptide substrates that can be degraded by ADEs, it is no surprise that many ADEs have been identified to have a number of other biological functions. This is particularly evident in two of the best known ADEs, NEP and ECE-1. Both enzymes have received considerable levels of interest for their roles in blood pressure regulation and represent promising therapeutic targets to combat cardiovascular disease. Both NEP and ECE-1 promote hypertension, however, each does so via different primary mechanism. To provide greater detail into each enzyme and a more complete view of their biological function the role of NEP and ECE-1 in blood pressure regulation will be considered below.

The primary mechanism by which NEP elicits blood pressure regulation is through natriuretic peptide cleavage (Figure 1.7). Natriuretic peptides (NPs) are separated into three major classes, atrial natriuretic peptide, B-type natriuretic peptide and C-type natriuretic peptide (ANP, BNP and CNP). All three NPs are small (between 22-32 amino acid residues in length) and possess a partially conserved cyclic core. ANP and BNP are found within arterial myocytes in mammalian hearts where they can be readily released into the bloodstream. CNP is expressed in endothelial cells of vascular tissues in addition to neuronal cells. While NPs exhibit variation in localisation each promotes vasodilation through binding to plasma membrane transmembrane receptors (Zois *et al.,* 2014).

Receptors for NPs are expressed in a wide range of tissues. Two guanylyl cyclase receptors for NPs, GC-A and GC-B, are responsible for cGMP generation. Protein kinase G then mediates downstream effects including calcium sequestration. The third NP receptor, NPR-C, does not possess guanylyl cyclase activity, instead it functions as a clearance receptor. Internalisation of NPR-C with a ligand leads to lysosomal hydrolysis of the ligand and NPR-C recycling (Potter, 2010).

The role of NEP's in NP degradation has led to it becoming a major drug target for heart failure. Several drugs for NEP inhibition have been developed and used to reduce hypertension. Currently one NEP inhibitor has been FDA approved for individuals with high cardiovascular disease risk. The inhibitor, sacubitril, is a pro-drug converted to sacubitrilat and functions as a dual inhibitor against NEP and ACE (Jhund and McMurray, 2016).

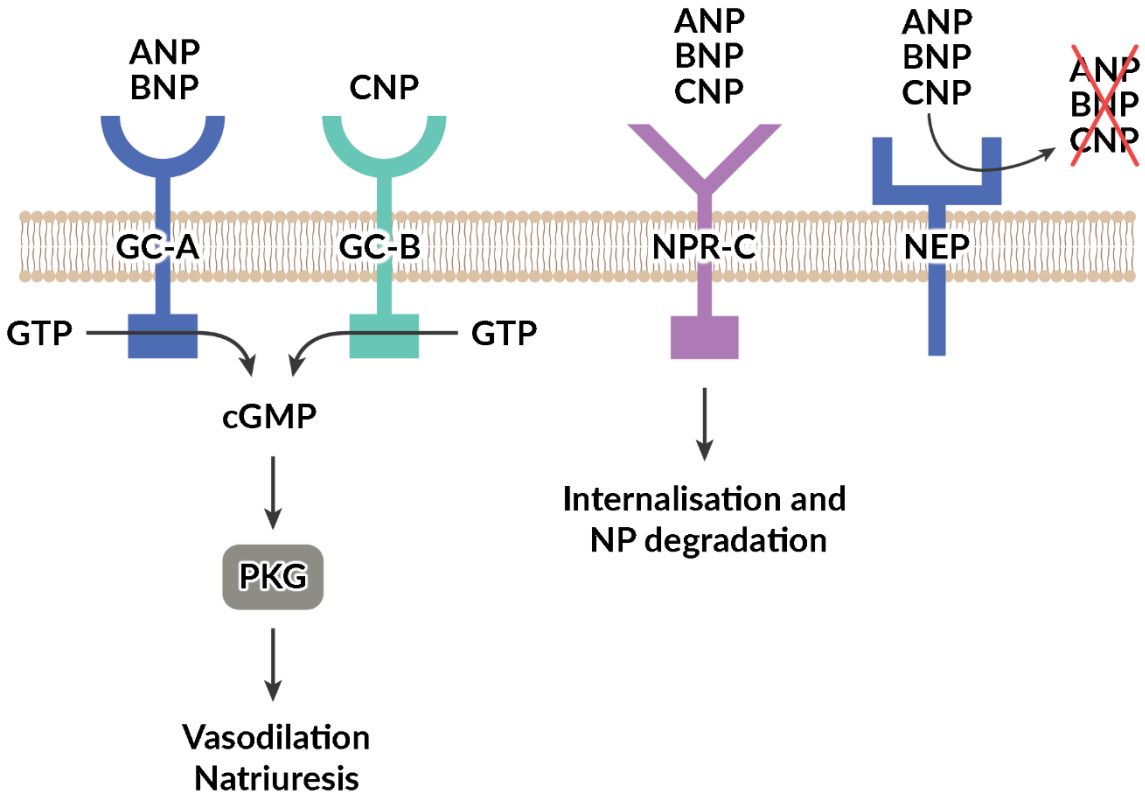


Figure 1.7, The role of NEP and natriuretic peptides in blood pressure regulation. Natriuretic peptides ANP, BNP and CNP interact with receptors GC-A and GC-B. These receptors then mediate an intracellular response through protein kinase G which leads to vasodilation and natriuresis. The third natriuretic peptide receptor, NPR-C, is responsible for peptide clearance. The transmembrane zinc metalloprotease, NEP, degrades natriuretic peptides leading to vasoconstriction.

The primary mode of blood pressure regulation for ECE-1 has been proposed to be through cleavage of big endothelins into endothelins (Figure 1.8). Three 21 amino acid ET isoforms have been characterised ET-1, ET-2 and ET-3. These ET's interact with receptors ET_A and ET_B (Masaki *et al.*, 1991; Haynes *et al.*, 1995). ET_A, predominantly located in vascular smooth muscle cells, is a G-protein coupled receptor responsive for mediating vasoconstriction. Upon binding to ET, ET_A activates the phospholipase C-inositol triphosphate pathway,

increasing the intracellular calcium concentration and leading to smooth muscle cell contraction (Miyauchi and Masaki, 1999). ET_B, located in both vascular smooth muscle cells and endothelial cells, has a contrasting role and is responsible for release of nitric oxide, ET clearance and ECE-1 inhibition. All of which elicit a vasodilatory response (Luscher and Barton, 2000).

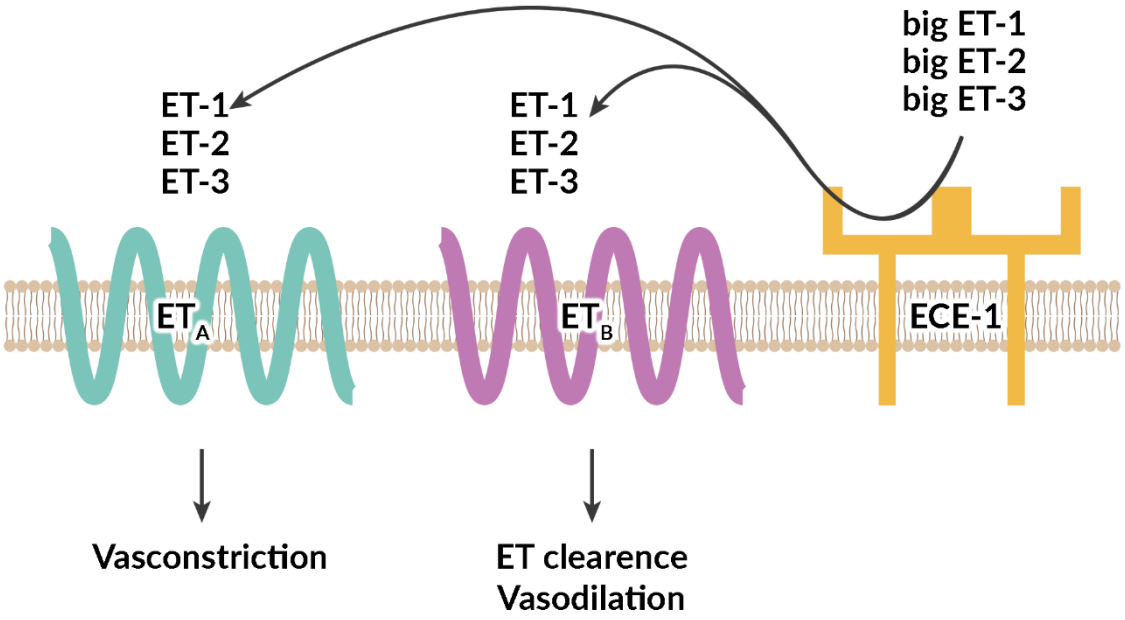


Figure 1.8, The role of ECE-1 and endothelins in blood pressure regulation. Big endothelin's ET-1, ET-2 and ET-3 are processed by transmembrane zinc metalloprotease, ECE-1. The resulting peptides can then interact with receptors ET_A and ET_B. The predominant response is vasoconstriction mediated through ET_A with ET clearance and some vasodilatory effects mediated by ET_B.

While ECE-1's role in ET production has been the primary focus of research, ECE-1 has also been revealed to exhibit protease activity on natriuretic peptides (Nakayama *et al.,* 2012). The physiological significance of this is not completely clear but it is likely that ECE-1, like NEP, has a role to play in natriuretic processing. The role of ECE-1 in blood pressure regulation has seen it become a drug target along with NEP and ACE. Proposed drugs are frequently designed as dual/triple inhibitors against two or three of the proteins (Mellin *et al.,* 2005; Kalk *et al.,* 2011; Seed *et al.,* 2012).

1.3 Aims of the PhD project

The role of ADE's in amyloid-β degradation is critical for brain homeostasis with imbalances strongly linked to Alzheimer's disease. While the potential to use ADEs as a therapeutic target has received some research attention, they have not been characterised in-depth. Further research into ADE's will provide a further base of knowledge of the processes underlying Alzheimer's disease and may prove to be invaluable for development of future therapies.

Two ADE's, NEP and ECE-1, represent high priority targets for greater research and characterisation due to their high potencies and their variation in subcellular localisation. This means together these proteins are highly efficient at degrading both extracellular and intracellular amyloid-β. NEP and ECE-1 also possess additional functions, most notably their role in blood pressure regulation. These additional functions further contribute to their priority for characterisation as any scientific advancements will have benefits across a number of research fields.

The specific aim of the Ph.D. project detailed in the following chapters was to characterise NEP and ECE-1 using a variety of biophysical techniques. Of these the primary method of characterisation would be through X-ray crystallography. Determination of high-resolution ADE structures, in both substrate-free and substrate-bound forms, would provide the basis for understanding how these proteins function and can fulfil the range of roles through degradation of a variety of substrates.

Chapter 2 – Structural biology and X-ray crystallography

2.1 Structural biology and X-ray crystallography

Life in its cellular and non-cellular forms is created from an immense number of macromolecules, each with a specific function and unique network of interactions. The field of structural biology investigates these fundamental macromolecules through the lens of structure. It is through structural determination that one can infer function and how these functions are mechanistically fulfilled. Moreover, characterisation of the individual macromolecules provides a means to understand the architecture of cells and multicellular organisms thus providing a clear foundation for many other biological sciences.

The bulk of structural biology research focusses on determination of protein macromolecule structures with a lesser focus on DNA/RNA. The most obvious issue with studying proteins is their size, commonly ranging from 1-20 nm in length with their covalently bound atoms 1-2 Å apart. The wavelength of radiation is approximately equal to its resolving power – the distance between two points that can be separated. Visible light (350-700 nm), even with the use of sophisticated microscopes, can only separate two particles approximately 500 nm making it ill-suited for the determination of atomic coordinates.

There are many techniques in the toolbox of a structural biologist including: mass spectrometry, small-angle X-ray scattering, circular dichroism, isothermal titration calorimetry and fluorescence resonance energy transfer. These can provide insights into the size, shape, fold and dynamics of a protein. However, if the primary aim of study is focused on determination of the precise atomic coordinates within macromolecules then one of three technique should be utilised. These are X-ray crystallography (XRC), nuclear magnetic resonance (NMR) spectroscopy and cryo-electron microscopy (cryo-EM).

XRC represents the 'gold standard' for structural determination. Since the first reported protein structure in 1958 (Kendrew *et al.,* 1958), XRC has been responsible for the dramatic rise in protein structure determination. To date, the technique has been used to determine over 130,000 structures, accounting for 90% of the total structures deposited into the Protein Data Bank (PDB). The importance of XRC in the development of structural biology cannot be overstated.

X-rays have wavelengths between 10 - 0.01 nm, which is similar to the distances between covalently bound atoms and far shorter than the wavelength of visible light. In XRC these

X-rays are focused on protein crystals, which contain a regular organisation of atoms, to produce X-ray diffraction. The amplitude of the diffracted X-rays can be determined from the recorded intensity of diffracted spots. This, in addition to the phase information, which can be estimated using a number of different methods, is used to determine precise atomic coordinates (Blow, 2004). Currently XRC represents a rapid and reliable method that is limited only by the requirement for proteins to be able to crystallise.

The second technique, NMR, accounts for approximately 8% of protein structures deposited in the PDB. The technique exposes nuclei to an external electromagnetic field which can be absorbed and emitted to produce a spectrum. Through analysis of this spectrum, information on atom composition and atom local environment can be gained. Whilst NMR has no requirement for protein crystals, it is limited by the size of protein that can be determined. Structures are commonly between 5-25kDa, but the upper limit is approximately 35kDa (Yu, 1999).

The third technique, cryo-EM, has seen a large increase in popularity in recent years, now contributing 2% of the total protein structures in the PDB. This technique measures the scattering of electrons produced by mono dispersed proteins randomly orientated on a grid. The scattered electrons from 1000's of protein molecules are then analysed and merged to recreate the protein structure. It has been suggested that we are on the cusp of a cryo-EM revolution. However, like NMR, cryo-EM suffers from limitations in protein size that can be determined. To date, the smallest protein determined is the 52-kDa streptavidin (Fan *et al.,* 2019), but structures are more commonly above 150kDa in size. Until recently, resolution has also presented a major limitation to the method but with recent advancements in equipment many atomic resolution structures are being reported (Renaud *et al.,* 2018).

Much of the work contained within this thesis is focused on the production protein for XRC. Once protein has been produced a routine collection of steps can then be performed in order to determine macromolecular structures (Figure 2.1). The remainder of this chapter explores the practical and theoretical aspects of these stages in XRC providing a basis for understanding later results chapters.

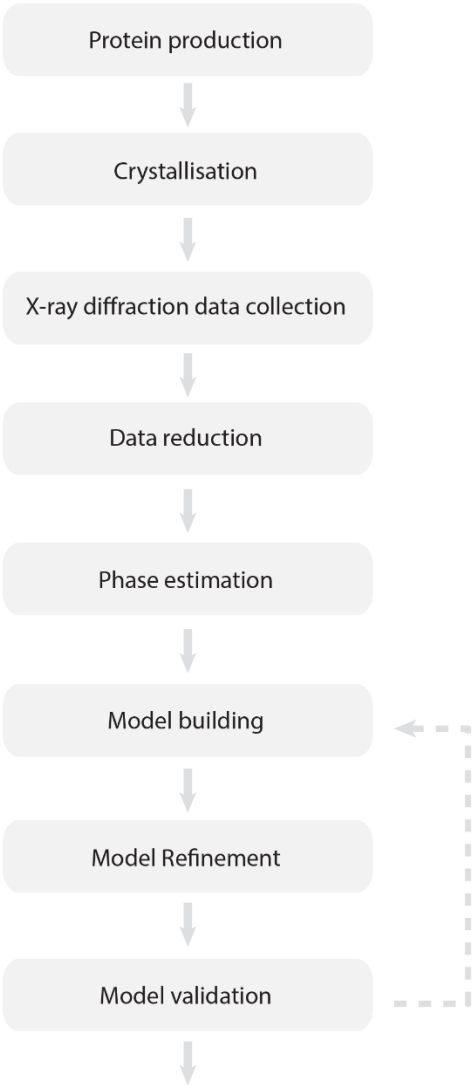


Figure 2.1, From protein to structure: a flow chart of X-ray crystallography methods. The stages involved in determination of a macromolecular structure using X-ray crystallography. The first three stages - protein production, crystallisation and data collection - are experimentally performed. Once data have been collected the following stages are computationally performed.

2.2 Crystallisation

To determine a protein structure using XRC it is necessary to obtain protein crystals. The process of crystallisation requires a large quantity of protein (>1 mg) and for that protein to be pure. Impurities may disrupt crystallisation or may indeed crystallise themselves. It is only after enough pure protein has been produced that crystallisation trials can begin.

Crystallisation trials are performed with the aim of identifying which, of a vast matrix of crystallisation conditions, can lead to crystal formation when mixed with concentrated protein solution. Trials can be conducted using several experimental techniques, the most common of which is vapour diffusion in the form of hanging or sitting drop (Figure 2.2a and b) (Dessau and Modis, 2011). In this method a reservoir is filled with reservoir solution and a crystallisation drop is created by mixing a ratio, often 1:1, of protein solution to reservoir solution. The reservoir and crystallisation drop are then sealed. As the crystallisation drop contains a lower concentration of precipitant compared with the reservoir solution, vapour diffusion and a net transfer of water from the crystallisation drop to the reservoir solution occurs. Over time, the concentration of precipitant and protein increase in the crystallisation drop until the protein's solubility limit is exceeded and the drop becomes supersaturated. At this point protein crystals or precipitant form and the protein concentration is reduced allowing the solubility limit to be re-established. This process can be represented in a phase diagram (Figure 2.2c).

If conditions for crystal formation are identified, even if crystals are small or diffract poorly, optimisation can be used to frequently improve size and diffraction quality. Optimisation of initial conditions through incremental adjustments of chemical components, such as salts, precipitants and pH can improve the quality of X-ray diffraction. Changes to the physical parameters, such as temperature, volume and the method of crystallisation may also improve diffraction (McPherson and Gavira, 2014; McPherson and Cudney, 2014).

It is the process of crystallisation that presents the largest hurdle in XRC. Despite the introduction of 1000's of crystallisation conditions and high-throughput screening, there are no guaranties that a given protein will ever yield crystals. However, there are several methods to make a protein more conducive to crystal formation. These include, removal of flexible regions or domains and glycosylation sites to reduce overall flexibility and increase the number of stable crystal contacts (Dale *et al.,* 2003; Chang *et al.,* 2007). Protein aggregation also presents an issue during crystallisation. This is especially true of transmembrane proteins, where the hydrophobic nature of transmembrane regions greatly increases the propensity for protein aggregation. If transmembrane domains are small and single pass it is often possible to remove them and determine the remaining protein structure.

The theory behind crystal formation is well documented (McPherson and Cudney, 2014) and can be separated into two stages: nucleation and growth. Nucleation is the process of forming an initial nucleus or crystalline solid. In this process protein molecules collide in orientations that form favourable contacts and become arranged in a regular crystalline pattern. This can only occur in protein solutions that are supersaturated or in the 'nucleation zone'. The formation and disassembly of nuclei arises frequently in a supersaturated solution. It is only after a critical volume is reached that the nucleus becomes stable and continues to grow. As solid nuclei increase in size the solution then becomes less saturated and the 'metastable zone' is reach. Here the second stage of crystal formation, growth, can continue to occur below the nucleation zone. This process is described as heterogeneous nucleation. In homogeneous nucleation, the supersaturated nucleation point is not reached, but crystal growth can occur through the introduction of externally produced nuclei in a process called seeding (McPherson and Gavira, 2014).

Crystallisation is driven by change in free energy (ΔG). Two components contribute to free energy, these are enthalpy (ΔH) and entropy (ΔS) (Equation 2.1). For crystallisation to occur the process must have a negative ΔG and therefore be energetically favourable.

$$\Delta G = \Delta H - T\Delta S \qquad \textit{Equation 2.1}$$

ΔH is regarded as relatively insignificant as the intermolecular bonds in a protein crystal contribute very little energy in formation. Therefore, the major driver of protein crystallisation is ΔS. This may seem unlikely, as protein in solution has a higher ΔS than protein in a solid state, however the ΔS of the solution must also be considered. In solution, protein is surrounded by ordered water molecules that form a hydration layer. When a protein transitions from in solution to a solid state these water molecules are released and there is a net increase in ΔS (Rupp, 2010). It is only through crystallisation trials that an appropriate condition can be identified were the ΔG is negative for crystal formation.

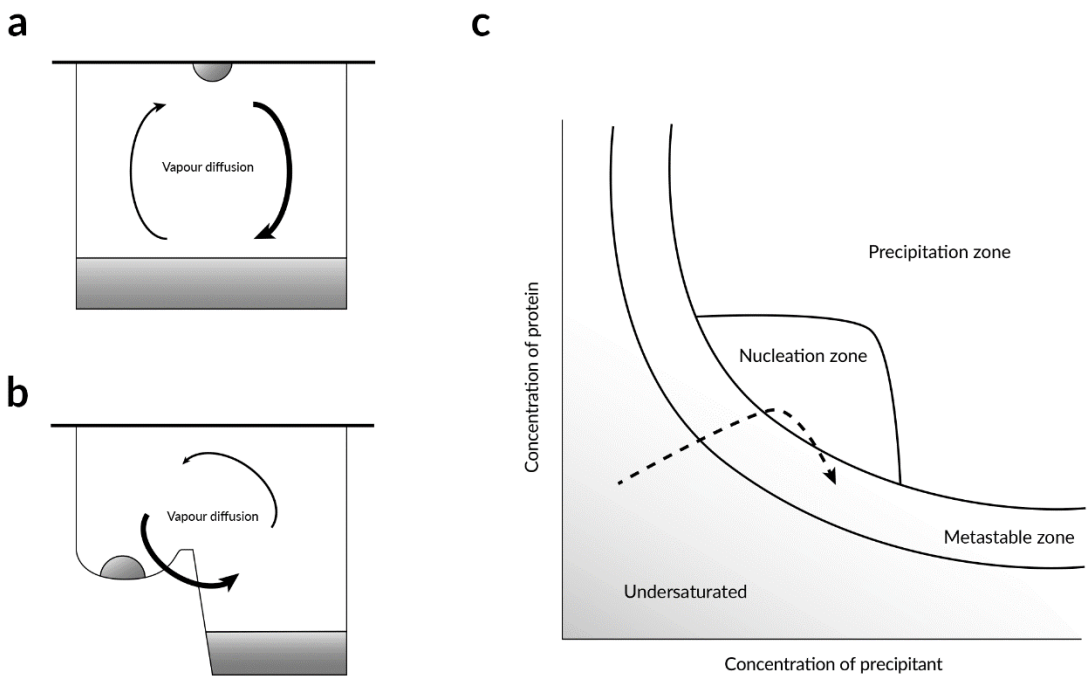


Figure 2.2. Protein crystallisation. (a) and (b) schematically represent the common experimental crystallisation experiments hanging and sitting drop vapour diffusion respectively. In these experiments a drop containing protein and crystallisation buffer is separate from a large reservoir of crystallisation buffer. Over time water diffuses from the protein containing drop into the reservoir due to the reservoir's higher precipitant concentration. (c) crystallisation diagram illustrating the path required for crystallisation of a protein. Initially the crystallisation drop will be undersaturated, but as vapour diffusion occurs the concentration of protein and precipitant rises until supersaturation is achieved. At this point crystal nucleation may occur and as crystals form protein concentration is reduced into a metastable zone where an equilibrium can be reached.

2.3 Crystal geometry

Crystals are 3-dimensional periodic assemblies of a homogeneous substance. In the case of a protein crystal, protein molecules form the identical repeating units that are periodically stacked to generate a complete crystal. Through studying the internal structure of different crystals, it has become clear that there are many ways that individual molecules can be organised. Information on this internal structure is critical for XRC to be used for protein structure determination. Two mathematical constructs – the asymmetric unit and unit cell - can be used to describe the internal geometry of a crystal.

The asymmetric unit is defined as the smallest unit of a crystal structure containing all the molecular information that can be translated and rotated using symmetry operators to generate the unit cell. Other types of symmetry can exist in a unit cell in addition to rotation, these are inversion and reflection. However, due to the chiral nature of protein molecules only rotational symmetry can exist in a protein crystal unit cell.

The unit cell is defined as the smallest unit that can produce the complete crystal lattice through translational operations alone. Given this property, the unit cell must be capable for tessellation, limiting the rotational symmetries to 0-, 2-, 3-, 4-, and 6-fold. Seven crystal systems exist which describe the symmetry contained within a unit cell (Table 1.1). It is this symmetry that places restrictions on both the axial lengths and interaxial angles of the unit cell which is drawn as a parallelepiped (Figure 2.3). Crystal system can be divided further into point groups, with each point group describing all the rotational operations that are experienced by the asymmetric unit within the unit cell (Rupp, 2010).

Further complexity may be present in the unit cell where, in addition to rotation, there may be some translational operations present between symmetry related molecules. If translation is present, it is given as a function of the unit cell length and the unit cell is said to contain a screw axis. The notation for screw axis is given as n_m, where 'n' indicates the rotation present and the translation is 'm/n' of the unit cell. For example, $P2_1$ has a rotation of 180^o for a translation of half the unit cell.

In addition to categorisation based on symmetry, unit cells can also be separated into 14 Bravais lattice groups based on the position of lattice points. Bravais lattices are constructed such that the axes of a unit cell axis are positioned parallel or perpendicular to the symmetry axes. In most cases a primitive unit cell can be constructed where lattice points are located at the 8 corners of the unit cell. However, it is sometimes necessary to construct larger unit cells in order to adhere to the convention of axis orientation. In these cases, unit cells can be generated with additional lattice points present inside, on the faces and at the ends (Blow, 2004).

The point group, screw axes and Bravais lattice can be combined into a space group. It is the space group that contains a complete description of the geometry contained within a crystal. 230 general space groups are available, however only 65 of these are available for

chiral molecules including proteins (Rupp, 2010). During an XRC experiment, X-rays are passed through a crystal and it is the space group that governs the produced diffraction pattern. Knowledge of the space group is required before the protein structure can be determined. The interaction of X-ray and crystals will be explored in greater detail below.

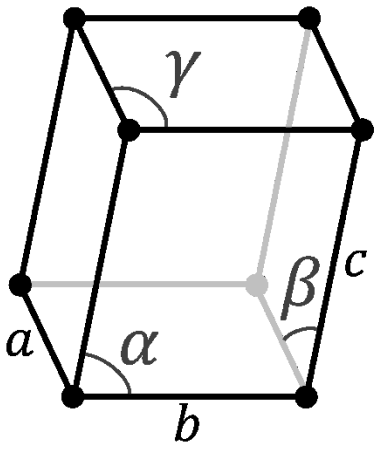


Figure 2.3. The unit cell. A single lattice points is shown at each edge. Axial lengths a, b and c and interaxial angles α, β and γ are restricted based on the symmetry operations experienced by the asymmetric unit.

Crystal system	***Interaxial angles***	***Axial lengths***	***Point groups***
Triclinic	α≠β≠γ≠90	a≠b≠c	1
Monoclinic	α=γ=90, β≠90	a≠b≠c	2
Orthorhombic	α=β=γ=90	a≠b≠c	222
Tetragonal	α=β=γ=90	a=b≠c	4, 422
Trigonal	α=β=90, γ=120	a=b≠c	3, 32
Hexagonal	α=β=90, γ=120	a=b≠c	6, 622
Cubic	α=β=γ=90	a=b=c	23, 432

Table 2.1. Crystal systems and unit cell restrictions. The seven crystal systems with the possible point groups are listed. Each of these crystal systems places restrictions on the unit cell interaxial angles 'α, β and γ' and axial lengths 'a, b and c'.

2.4 X-ray diffraction

The process of diffraction occurs when electromagnetic radiation encounters an obstacle and is scattered in directions other than the original one. In XRC, X-rays interact with electrons causing them to adopt a higher energy state. Upon returning to a lower energy state, radiation of the same wavelength is emitted in a random direction after a phase lag (Brill, 1968). X-ray diffraction by a single protein molecule is incredibly weak and cannot be used to determine the protein structure. In contrast, the diffraction of a protein crystals is far stronger due to constructive interference. From the observed diffraction pattern is possible to determine the crystal geometry in addition to the electron density distribution of the unit cell. It is this information that allows determination of a protein structure.

The theory of X-ray diffraction in 3-dimensional crystals is well documented (Rupp, 2010), but is it often more simply explained in 2-dimensions. Let us consider passing a parallel beam of monochromatic X-rays through a perpendicular 2-dimensional plane of regularly space atoms. The result would be that X-rays simultaneously reach each atom in the plane before being scattered in all directions. The majority of diffracted waves will be out of phase and thus cancelled out due to destructive interference. However, in some directions, diffracted waves from adjacent particles will vary in phase by a whole number of wavelengths. It is at these specific directions that constructive interference can occur. If the diffracted waves were recorded on a 2-dimensional detector then a regular arrangement of spots would be present. In XRC it is the position of spots and their intensity that can be used to determine the atomic structure that is responsible for the observed diffraction.

As atomic planes increase in complexity, more than one type of regularly spaced atoms can be present on the 2-dimensional plane. Each atom type would contribute a component of the total scattering in directions that allow for constructive interference. In this case the waves of each component will not be in phase hence when they interfere the resulting wave

will be the same wavelength but will vary in amplitude and phase to the incident waves (Blow, 2004).

These same principles apply to a 3-dimensional lattice which can be thought of as a stack of parallel 2-dimension planes. In a lattice however, all diffracted X-rays for a given plane must be in phase with those from all other parallel planes for constructive interference to occur Bragg's law (Bragg and Bragg, 1913) (Equation 2.2) describes the angle of incidence (θ) required for which X-rays of wavelength (λ) can be diffracted by parallel planes with spacing (d) in phase to produce constructive interference (Figure 2.4).

$$n\lambda = 2dsin\theta \qquad \textit{Equation 2.2}$$

It is clear that for a given plane constructive interference will only occur in some crystal orientations relative to the incident beam (angles of θ). Therefore, at a single angle of incidence the diffracted X-rays provide a fraction of the total reflections for a crystal. As the crystal is rotated in the beam, the angle of incidence changes causing different parallel planes to diffract waves in phase and results in different spots to become visible on a detector (Figure 2.5).

Planes will intersect the unit cell at different positions and once a full dataset is collected information on every scattering component will be known. Each plane corresponds to a distinct spot or reflection and can be assigned Miller indices '*h, k* and *l*' describing the position that the lattice plane intersects the unit cell. In some instances, the symmetry present within a unit cell causes planes that satisfy Bragg's law to be absent. These are named systematic absences and are a result of destructive interference from symmetry molecules. Systematic absences can be used during indexing to improve point group prediction of a given protein crystal (Rupp, 2010).

A given reflection can be described by resolution in addition to its Miller indices. A high-resolution reflection will have a higher scattering angle but will be less intense. The higher angle is caused by its lower interplanar spacing, which cause higher resolution spots contain more information about the electron density distribution within the unit cell. During XRC

experiments collection of high-resolution reflections is desired to gain the most precise atomic coordinates possible.

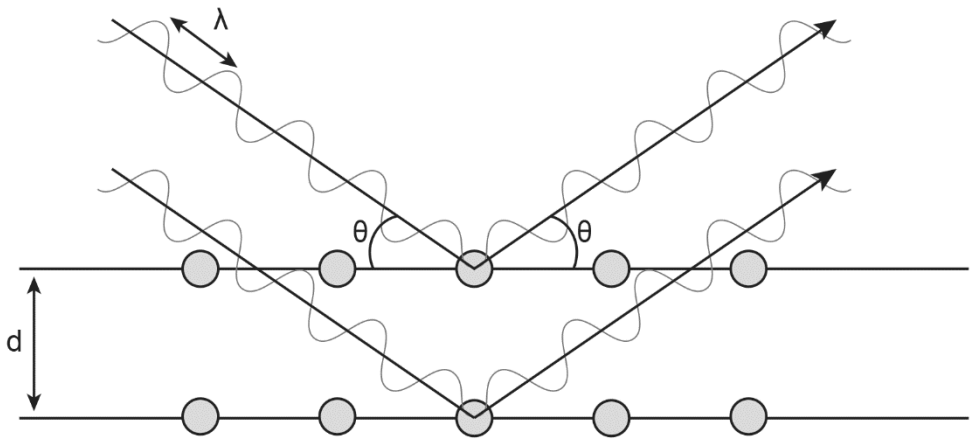


Figure 2.4, Bragg's law. A diffraction spot will only be detected when Bragg's law is satisfied for the angle of incidence (θ) the wavelength (λ) and the distance between parallel planes (d).

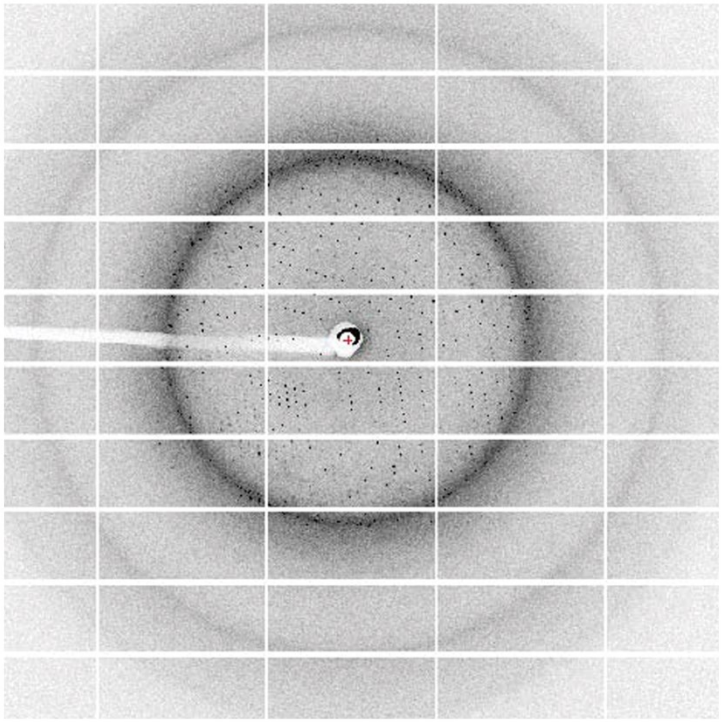


Figure 2.5, Protein crystal X-ray diffraction. The image shows the diffraction spots that are the result of crystallographic planes that satisfy Bragg's law at this orientation of crystal in the X-ray beam. Spots further from

the centre are higher resolution reflections. White grating pattern in the image indicates the edge of each detector module.

2.5 X-ray diffraction data collection

The aim of data collection in XRC is to accurately determine the positions and intensities of reflections from planes that satisfy Bragg's law to the highest resolution possible. In order to do this, there are several physical requirements: an X-ray source, a detector, a goniometer and finally a protein crystal.

X-ray sources can be classed as laboratory generators or synchrotrons. Laboratory generators frequently use a rotating-anode that is bombarded by electrons to generate X-rays, and whilst these generators provide intense X-ray beams, they are substantially weaker than synchrotrons. As such, data collection is much slower with an increased likelihood of detrimental radiation damage (Rupp, 2010).

At present, the majority of XRC data is collected at a synchrotron because of the collection speed and the tunability of individual beamlines (Helliwell, 2017). Synchrotrons are cyclic particle accelerators that generate X-rays using the acceleration and curvature of electrons to produce electromagnetic radiation. Briefly, an electron gun generates electrons by feeding a high voltage through a cathode. The produced electrons are accelerated, by a linear accelerator, before entering a booster ring for a further increase in energy. Once at the required energy, electrons are fed into the large storage ring where radio frequency generators are used to maintain the electron energy levels. In the storage ring electrons are circulated through straight sections connected by bending magnets. These magnets produce radiation by imposing a change in the direction of circulating electrons. The radiation can be tailored in wavelength and transmission in an optics hutch to facilitate the specific experiments taking place at a given beam line. Modern synchrotrons also contain insertion devices: undulators and wigglers. These devices are inserted into straight sections of the storage ring and an alternating magnetic field results in electron oscillation that emits higher energy radiation (Wiedemann, 2015).

Detectors of diffracted X-rays have developed greatly, transitioning from photographic film to electronic area detectors to image plates to charged-coupled devices and now pixel

detectors. Each iteration of detector promises improved signal to noise with faster collection and less dead time (Dauter and Wlodawer, 2016). The latest instalment of pixel detectors, Dectris Eiger2 X 16M, allows for the collection of a data set in minutes. The speed of modern data collection frequently means that the majority of time spent on a modern beamline is discerning the data collection strategy to be used rather than the data collection itself.

As previously discussed, to determine a protein crystal structure, diffraction data must be collected through multiple angles of the crystal. To achieve this, crystals are mounted on pins which, in turn, are mounted on goniometers. A goniometer functions to rotate a crystal, usually along a single Φ axis, during data collection. Modern goniometers exhibit many additional axes of rotation including the ω and κ axis as well as translational axis X, Y and Z. These allow for improved crystal positioning and the ability to undergo multiple rotational axis during data collection that limits radiation damage (Rupp, 2010).

The precision of modern goniometers allows for rapid data collection at small rotational increments (~0.1°). Collection of diffraction images from these small rotations referred to as 'fine slicing' which contrasts to 'wide slicing' where larger ~1° rotations are used. Using smaller rotational increments to capture the diffraction image has two main benefits. First, it removes the likelihood that overlapping reflections will be generated on a single image. Second, each reflection will be partially observed over a number of images and the total can be more accurately estimated while improving signal to noise (Dauter, 2017).

Clear advancements in X-ray source, detectors and goniometer have allowed for collection of higher resolution data in faster times. The final physical requirement for X-ray diffraction is the protein crystal itself. Whilst it is not possible to technologically improve a protein crystal, advancements have been made in the way crystals are handled. During data collection the high-intensity X-rays cause radiation damage to protein crystals. In order to limit the damage, cryocooling can be used. Protein crystals will be flash-cooled and stored in liquid nitrogen (63 – 77.2 K). For data collection to proceed, crystals are transferred from a liquid nitrogen container into the beam path. In this position a cryostream of cooled nitrogen gas is flowed over the crystal.

The process of flash-cooling, while important for limiting radiation damage, may have detrimental effects through the formation of ice crystals within the protein crystal. Not only

will this lead to potential disruption of unit cell stacking but it can lead to characteristic ice rings on collected diffraction images. These rings saturate the detector making identification of diffraction spots unachievable. To overcome ice formation the speed of cooling must be kept to a minimum to produce amorphous ice instead of crystalline ice. Cryoprotects, such as glycerol and poly-ethylene glycol, are also frequently used before flash-cooling to replace water content of crystals reducing the likelihood of crystalline ice formation.

Aside from the physical requirements for collection several parameters must be considered to optimise the collection process. These include the exposure time, transmission, beam size, wavelength, number of images to collect and distance to the detector. It is through adjustments of these parameters that collection of the highest resolution data while limiting the radiation damage to the crystal. Modern synchrotrons have pipelines that feed test shot data into software such as EDNA (Incardona *et al.,* 2009) to provide estimates for the best data collection strategy.

2.6 Data processing

2.6.1 Reduction

Once X-ray diffraction data has been collected it can be processed. There are four components to data processing, these are spot finding and indexing, integration, scaling and merging.

Spot finding and indexing provide estimates of unit cell dimensions and indications of the crystallographic symmetry. These processes are automated by programmes including MOSFLM (Battye *et al.,* 2011), XDS (Kabsch, 2010) and DIALS (Waterman *et al.,* 2016) and hence are usually referred to as autoindexing. First, spots are distinguished from background radiation and their position on the 2-dimensional detector is recorded. Each spot will be assigned Miller indices '*h, k* and *l*' which describes the lattice plane that has generated that specific reflection. Initially autoindexing uses a primitive triclinic cell (P1) as all diffraction for a chiral molecule can be assigned to group. Once indexed it is possible to identify if higher orders of crystal symmetry are present. This is done by scoring the data on how well they fit different point groups and Bravais lattices based on unit cell restraints (Powell, 2017).

MOSFLM, XDS and DIALS also record the intensity of each spot in the integration process. Intensities can be calculated by summation or profile fitting. Summation involves adding all pixel values that form a given spot, while profile fitting aims to model the profile of a spot as a smooth curve and then calculates the area underneath. Profile fitting will also, in the case of fine slicing, estimated the curve by combining the values pixels over to course of several images where partial reflections are observed. The calculated intensity values from either method are then subtracted from background radiation (Powell, 2017).

Intensities of the integrated spots will be non-uniformed. This is due to variation in a number of factors during data collection. These factors include X-ray beam intensity, crystal radiation damage, anisotropic absorption of the crystal and movement of position in the beam. As a result, the recorded intensities need to be adjusted or 'scaled' in order to account for these variations seen in data collection. Scaled data can then be merged by averaging the observed intensity of each reflection across a number of images in order to obtain a value that is closer to the true value of the reflection. Aimless (Evans and Murshudov, 2013) is the most widely used programme that can both scale and merge data.

It is important to note that much of these data reduction processes at modern synchrotrons have been automated in pipelines that included xia2 dials, xia2 3dii (Winter, 2010), autoPROC (Fodje *et al.,* 2014) and FastDP (Winter and McAuley, 2011), such that minutes after data collection results from several pipelines will be provided.

2.6.2 Quality of data

After merging, the quality of reflection data is assessed. Reflections are organised into bins or resolution shells which contain reflections within a certain resolution range. This is necessary as reflections at a higher resolution are weaker and must be analysed independently of other reflections to accurately determine a resolution cut-off. Several statistics are used to assess the quality of each resolution bin and the data as a whole. A resolution cut-off can be determined from this information and applied to the data. This cut-off will exclude noise from the data that can otherwise negatively affect the calculated electron maps.

R-factors have been the most common measures of data quality. R_{merge} (Equation 2.3) and R_{meas} (Equation 2.4) are measures of the uncertainty present in the data prior to merging.

Specifically, the discrepancy of each intensities (I) for a given reflection (*hkl*) compared with the mean of that reflection where 'N' is number of symmetry related partners. Modern collection techniques frequently use high multiplicity strategies because the more a reflection is observed the more precise the intensity estimation. This is problematic when using R_{merge} and R_{meas} given that they both increase with multiplicity and therefore provide a poor quality indicator (Weiss and Hilgenfeld, 1997).

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^{N} |I_{i\,(hkl)} - \bar{I}_{(hkl)}|}{\sum_{hkl} \sum_{i=1}^{N} I_{i\,(hkl)}} \qquad \textit{Equation 2.3}$$

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{N}{N-1}} \sum_{i=1}^{N} |I_{i\,(hkl)} - \bar{I}_{(hkl)}|}{\sum_{hkl} \sum_{i=1}^{N} I_{i\,(hkl)}} \qquad \textit{Equation 2.4}$$

R_{pim} (Equation 2.5) on the other hand, describes the precision of the averaged intensity measurement while also taking into account the multiplicity of the data. So unlike R_{merge} and R_{meas}, it can be used to provide an accurate estimation of resolution cut-off (Weiss, 2001).

$$R_{pim} = \frac{\sum_{hkl} \sqrt{\frac{1}{N-1}} \sum_{i=1}^{N} |I_{i\,(hkl)} - \bar{I}_{(hkl)}|}{\sum_{hkl} \sum_{i=1}^{N} I_{i\,(hkl)}} \qquad \textit{Equation 2.5}$$

In addition to R-factors, there are several other measures of data quality that can be used to determine a data resolution cut-off. The signal to noise measure (Equation 2.6) is calculated for each reflection (*hkl*) and the average intensity over symmetry mates (I_{hkl}) is divided by estimated error $\theta(I_{(hkl)})$. Commonly an $\langle |I|/\sigma(I) \rangle$ = >1 is used as a cut-off in the

highest resolution shell as intensities remain significantly higher than the noise (Evans and Murshudov, 2013).

$$\langle |I|/\sigma(I) \rangle = \frac{1}{N} \sum_{hkl}^{N} \frac{|I_{(hkl)}|}{\sigma(I_{(hkl)})} \qquad \textit{Equation 2.6}$$

A more recent statistic that is described as a better indicator of data quality is $CC_{1/2}$ (Equation 2.7). This statistic first separates the recorded intensity measurements into separate resolution bins and then multiple measurements of the same reflection are randomly separated into two bins of equal size. A Pearson correlation coefficient (CC) is then determined for agreement between the two bins of measurements (Karplus and Diederichs, 2012; Wang *et al.,* 2017). At low resolutions $CC_{1/2}$ is close to 100%. Higher resolution bins will have a lower $CC_{1/2}$ as the data will contain more noise. Commonly data cut-off will be applied at a resolution where $CC_{1/2}$ falls below 30% (Evans, 2006; Karplus and Diederichs, 2012) but there are suggestions that it should go to 10% in the outer shell (Karplus and Diederichs, 2015). However, there are examples where including data below the a $CC_{1/2}$ of 27% does not provide additional structural information (Evans and Murshudov, 2013).

$$CC_{1/2} = \frac{\sum (a_i - \langle \bar{a} \rangle)(b_i - \langle \bar{b} \rangle)}{\sqrt{\sum (a_i - \langle \bar{a} \rangle)^2 \sum (b_i - \langle \bar{b} \rangle)^2}} \qquad \textit{Equation 2.7}$$

To determine a resolution limit for a data set it is necessary to review multiple data quality statistics. There is a balance to be had when it comes to determining a resolution cut-off. It is possible to remove noisy high-resolution data and in doing so improve data quality

statistics. However, removal of weak reflections is throwing out valuable information that ultimately may improve the structure. These data reduction statistics, with data collection parameters and unit cell parameters, are usually included in any XRC reported as 'crystallographic data statistics'.

2.7 Structural determination

Once diffraction data are collected and have been processed in a manner that yields the most information, the election density equation (Equation 2.8) can then be used to determine the 3-dimensional distribution of electron density for all atoms [p(xyz)] within the unit cell. However, in order to determine electron density, it is necessary to determine the structure factors.

$$p(x\ y\ z) = \frac{1}{v}\sum_{hkl}|F(hkl)|e^{-2\pi i[hx+kx+lz-\varphi(hkl)]} \qquad \textit{Equation 2.8}$$

The structure factors for each reflection, with a given Miller index (*hkl*), have both amplitude [F(*hkl*)] and phase [φ(*hkl*)] components. Using the Fourier transform [$e^{-2\pi i(hx+kx+lz)}$] these structure factors can be converted into an electron density map. The amplitude for a given reflection is proportional to the square root of intensity which is measured during X-ray diffraction data collection. However, the phase component of the diffracted waves is lost during data collection and must be estimated. This absence of phase information is often referred to as the 'phase problem' (Taylor, 2003) and methods to solve a structure through phase estimation or otherwise are discussed below.

2.7.1 Patterson methods

The structure of small molecules can be determined without phase information using the Patterson method (Patterson, 1935). Here the phase of each reflection is given the value of

zero. Using just the recorded intensities a Patterson map can be generated that reveals the interatomic distances of atoms within the unit cell which can be used to determine real space atomic positions. However, as the number of atoms (N) within the unit cell increases as does the number of peaks present on a Patterson map by N(N-1). As such, the method can only be used for structures where N < 50 (Cowtan, 2003).

2.7.2 Direct methods

Direct methods are named because phases are estimated directly from the diffraction pattern. Structures of small to intermediately (N<100) (Usón and Sheldrick, 1999) sized molecules may be solved with the absence of phase information, like via the Patterson method, if X-ray diffraction data is collected to a high resolution (~1.2 Å or better). High-angle diffraction spots give information about features of the unit cell that can be used to reconstruct the phases via the mathematical relationship between phases and structure factors. This method relies on prior knowledge of electron density – electron density cannot be negative, and it is generally organised into spheres around atomic peaks.

Generally, the structure factors are calculated for a set of strong reflections. Once a set of structure factors are known this information, due to the phase relations between structure factors, can be used to quickly estimate the remaining phase information and determine the electron density in the unit cell (Rupp, 2010).

2.7.3 Experimental methods

The two methods described above can be used for molecules with small to intermediate numbers of atoms. However, to determine a protein structure, which contains many thousands of atoms, experimental methods can be utilised.

A common experimental method is multiple isomorphous replacement (MIR) (Perutz, 1956). This method compares the diffraction pattern produced from native crystals to crystals where known contents have been added to a unit cell. The added contents, commonly heavy metal atoms, function as scattering centres which produce a detectable change in the intensity of some diffraction spots. By comparing the diffraction patterns, it should be possible to identify if the presence of the heavy metal has led to constructive interference (more intense spot) or destructive interference (less intense spot). With this in mind, it is critical that the added heavy metal atoms should adopt the same position in all unit cells

across the entirety of the protein crystal and should not alter the position of the other atoms in the unit cell.

Comparison to just one type of heavy metal-soaked crystal results in ambiguity as to whether unit cell atoms lead or lag the heavy metal atoms. Often several crystals with different heavy atoms bound are used to resolve this uncertainty. This allows for an estimation of the difference in phase between the heavy metal atoms phase and the remaining unit cell atoms to be generated. The position of these heavy metal atoms can then be calculated via the Patterson method in order to calculate their respective phase. Once the structure factors for the heavy metal atoms have been determined it is possible to approximate the phase from the remaining scattered X-rays (Cowtan, 2003).

Multiwavelength anomalous dispersion (MAD) is the second major experimental method used for phasing (Hendrickson, 1991). This method exploits the tuneable wavelengths of modern synchrotrons and the absorption edge for specific atoms. An absorption edge is the energy at which there is a dramatic increase in the absorption coefficient of X-rays by an element. As such, collecting several X-ray diffraction datasets using wavelengths around the absorption edge will generate different scattering patterns and the contribution from the anomalously scattering atom can be identified. Selenomethionine is frequently used to replace methionine for the purpose of MAD. Once the small number of anomalously scattering atoms have been identified it is then possible to estimate the phase of the remaining atoms like in MIR (Cowtan, 2003).

2.7.4 Molecular replacement

The previous methods that have been described are incredibly powerful when no structural information is available for a protein of interest. However, when structural information is known, such as when the protein of interest is similar to another protein either in function or sequence identity, then molecular replacement can be used. The aim of molecular replacement is to position a probe into the unit cell in a way that would produce hypothetical diffraction data identical to that of the experimental diffraction data.

Before molecular replacement can be used it is necessary to identify the symmetry and unit cell dimensions that exist within your protein crystal. Estimations of these parameters can be gained during data reduction, specifically the indexing step. The number of molecules

within an asymmetric unit must also be known and can be estimated using Matthews coefficient (Matthews, 1968). In addition to information regarding the crystal geometry an appropriate probe must also be selected. If a protein has high sequence similarity to another protein, then that entire protein may be used as the probe. If no close homologues are available, the choice of probe may require careful thought and involve using smaller fragments of proteins. Once an appropriate probe is selected several programmes such as Phaser (McCoy *et al.,* 2007) or MOLREP (Vagin and Teplyakov, 2010) are available for molecular replacement.

The process of molecular replacement can be separated into four stages:

1. A rotational search for the probe so that it is parallel to the displaced molecule
2. A translation search for the probe so that it is in the position of the displaced molecule.
3. The generation of a hypothetical model of the protein of interest based on the search model
4. Refinement of the model using diffraction data collected for the protein of interest

A given macromolecule will have a set of intra-atomic (self-molecule) and inter-atomic (symmetry molecules) vectors. In general, the intra-atomic vectors will be shorter and can be separated from the inter-atomic vectors. A Patterson map of the intra-atomic vectors is used to validate the initial rotational search. When the model is rotated the intra-atomic vectors rotate by the same angle. It is possible to rotate the probe until the intra-atomic vectors present on a Patterson map align with that of the experimental data (Evans and McCoy, 2008). Once the correct rotational orientation has been determined, the probe can be translated in the unit cell. Again, the Patterson map, now with inter-atomic vectors included, can be used to validate when the probe is in the correct position. By separating these searches, the problem becomes one that is possible to solve computationally.

In addition to Patterson methods for validation of a solution, maximum-likelihood is another method to check possible structural solutions. This method uses a statistical approach,

measuring the probability that the observed diffraction would occur given the position of the probe in the unit cell. Like the Patterson method, first rotation of the model is assessed before translation. Unlike the Patterson method, errors are incorporated for both the experiment and model which gives more robust solutions when compared to the Patterson methods. (Read, 2001)

In recent years, advancements in molecular replacement have enabled phase determination of large structures through the use of very small probes often with sizes in the region of ten amino acids. Programmes such as Fragon (Jenkins and IUCr, 2018) and Arcimboldo (Sammito *et al.,* 2014) are able to achieve this using prior knowledge of the secondary structure features of protein macromolecules. For example, α-helices have a very distinct electron density shape which can be exploited to model helical peptide fragments. It is also possible to model β-strands however the less identifiable density limits the method. After initial phases are estimated, model building programmes such as Buccaneer (Cowtan, 2006) or ARP/wARP (Perrakis *et al.,* 2001) can then be used to trace the main chain and complete the rest of the structure. These methods require high resolution data (<1.7 Å) for the electron density to be clear enough to model these small initial small probes.

As well as advancements in molecular replacement programmes, the number of determined protein structures has led to the development in the generation of probes. *Ab initio* modelling exploits the modular nature of protein structures. If the sequence of a protein of interest is known but has no obvious structural homologues the protein can be separated into shorter segments. The secondary structure of these segments can be predicted by searching large databases for similar sequences. The segments can then be linked, and energy minimised, in order to create a three-dimensional model for the protein structure. Such programmes that are able to build these models include Ample (Bibby *et al.,* 2012) and Phyre2 (Kelley *et al.,* 2015).

In addition to the methods of molecular replacement described above several molecular replacement pipelines, including Balbes (Long *et al.,* 2008) and MoRDa (Vagin and Lebedev, 2015), are also available. Molecular replacement pipelines require only a protein sequence and integrated diffraction data and can generate appropriate probes before attempting to determine the phases for the structure. The programmes and pipelines currently available usually make molecular replacement a trivial task. It is only in some circumstances, where

proteins have no homologues or may undergo large conformational changes, that molecular replacement may present a greater problem.

2.8 Refinement and validation

Once the phases for the macromolecular structure have been estimated and an initial model of the structure has been constructed, model refinement can take place. This process involves alteration of the modelled structure with the aim to improve the agreement with the observed structure factors while satisfying chemical requirements for bond lengths, angles and atomic interactions.

In structural refinement the calculated structure factors from the model are compared with the observed structure factors from the protein crystal. This can be visuallised using electron density maps 2Fo-Fc and Fo-Fc where 'Fo' refers to the observed structure factors and 'Fc' to the calculated structure factors based on the model. The Fo-Fc map or 'difference' map shows differences between the observed and calculated structure factors. Often this map provides insight as to whether the model needs additional atoms to account for an area of electron density or whether atoms need to be removed due to an absence of electron density.

Real space refinement programmes such as REFMAC5 (Murshudov *et al.,* 2011) and Phenix (Adams *et al.,* 2010) use successive cycles to make adjustments to the entire structure, improving model agreement with observed data while balancing geometric restraints. The weighing of geometric restraints can be manually adjusted. It is common practice to place a higher weighting on these restraints when refining lower resolution data as precise atomic coordinates are less clear from the data.

Manually editing the protein model with programmes such as Coot (Emsley *et al.,* 2010) allows addition and repositioning of atoms to better fit into electron density. It is often necessary to reposition atoms in order overcome an energy minimum that would not be overcome in real space refinement programmes. Structural refinement is done over several rounds of real space refinement and manual editing. After each round, the phases are

recalculated for the new model and an improved electron density map will be generated. Together these can be used to make incremental improvements to the structure.

During the refinement process it is essential to validate the structure to confirm that the refinement stratagy is indeed improving the model. There are several metrics for structural validation. The R-factor or reliability factor (Equation 2.9) is a measure in the correlation between the generated crystallographic model and the experimental X-ray diffraction data (Rupp, 2010). In real space refinement programmes this is the correlation between calculated structure factors to the observed structure factors (F_{cal} and F_{obs}). R = 0 when there is complete agreement between the calculated and observed structure factors. R ~ 0.59 for a completely random arrangement of atoms within the unit cell (Karplus and Diederichs, 2012). In practice R-factor values never reach zero as all data collected are imperfect and are limited by resolution.

$$R = \frac{\sum_{hkl}||F_{obs}|-|F_{calc}||}{\sum_{hkl}|F_{obs}|} \qquad \textit{Equation 2.9}$$

It is possible the reduce the R-factor by adding atoms into noise but this will not improve the quality of the model. To combat this, the statistic 'Rfree' is used during refinement to prevent model bias or overfitting into the observered data. Rfree is calculated by excluding a small number of reflections (frequently 1000 unique reflections or up to 5% of total unique reflections), evenly selected across resolution bins, from the structure calculation. The agreement of these reflections to the calculated model is used to provide a measure of the overall fit while removing bias.

In addition to the R-factors, which describe how well the model fits the data, chemical restraints must also be considered throughout refinement. The geometry of atoms within amino acids, and amino acids within polypeptide chains, have been determined in crystal structures with a high degree of accuracy. The geometry of these varies only slightly in different protein structures. As such, this prior knowledge allows restraints to be imposed onto a protein structure during refinement. These are particularly useful for lower resolution

data where exact atomic position is not known and restraints help to predict the most likely position.

Bond lengths and angles have been studied in small molecule crystal structures and variation is relatively low despite changes in the local environment (Engh *et al.,* 1991). Root-mean-square deviation (RMSD) is used as a measure of deviation between the ideal angle and lengths and those that are modelled. For bond lengths ~0.02 Å is considered good quality but above ~0.03 Å may indicate issues in the model (Wlodawer *et al.,* 2008) and for bond angles < 2° is desirable (Jaskolski *et al.,* 2007).

Another common measure of geometry used during refinement are the Ramachandran restraints. These restraints are a set of dihedral angles that can be adopted for an amino acid given the position of the amino acid previously specified. As atoms in a peptide bond lie in a plane, the main chain and C_β atoms in two adjoining residues can be described by two angles – Ψ and Φ (Ramachandran *et al.,* 1963) (Figure 2.6). Only some of these angles are energetically possible. The effect of side chains on the allowed Ramachandran angles is small with the exception of amino acids glycine and proline. Glycine posesses a side chain consisting of a single hydrogen atom. The absence of a C_β atom allows for a larger range of angles to be adopted given the flexibility of the residue. In contrast, proline has a smaller range of angles that can be adopted due to the side chain connecting the C_α atom and the nitrogen of the amino group. The rigid nature of proline residues also impacts the preceeding residue to again limit the number of energetically allowed angles. Based on the limited Ψ and Φ angles a Ramachandran plot can be constructed (Figure 2.7). It should be noted that Ramachandran assumes fixed bond lengths and bond angles. Sometimes the main chain or side chain interactions of a amino acid cause the residue to adopt angles that fall outside of the Ramachandran allowed region. This occurs infrequently, and depending on protein size may happen up to a few times across the entire structure.

Several other measures should be considered during refinement including rotamer outliers and clashscore. After a structure is appropriately refined and validated it can be deposited into the PDB. Once deposited, structures can be accessed and download by users. The

quality of deposited structures can be quickly assessed with the 'percentage ranks' statistics associated with each structure.

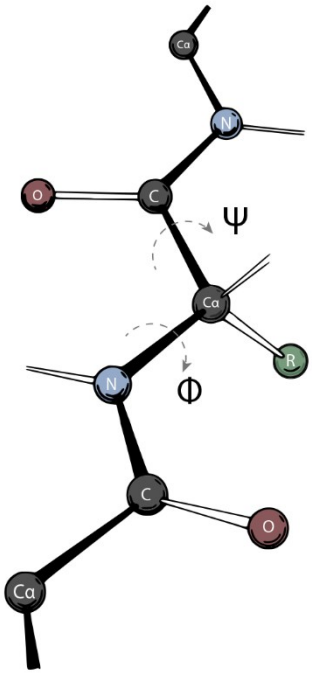


Figure 2.6. Ramachandran restraints. The atoms of a peptide bond exist in a plane, the remaining bonds within an amino acid can be described with two dihedral angles Ψ and Φ. Only some angles of Ψ and Φ are energetically available.

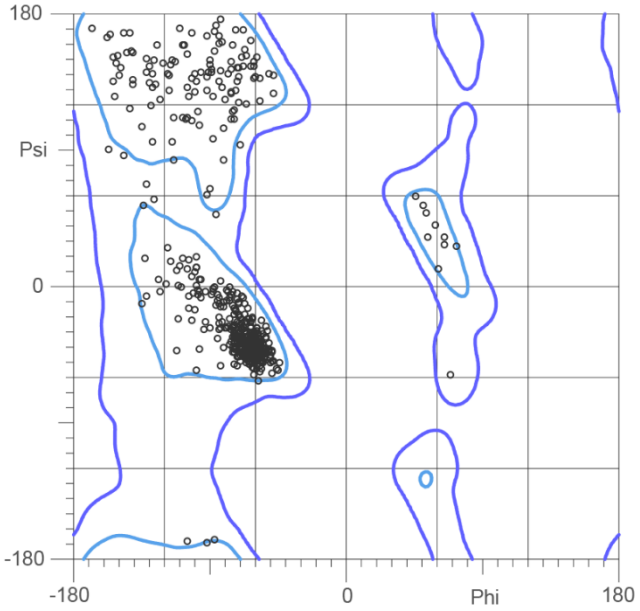


Figure 2.7, Ramachandran plot. The Ψ and Φ angles for a given amino acid can be plotted onto a Ramachandran plot. Areas of the plot will be favoured, allowed or unfavoured based on Ψ and Φ angles restraints. Amino acid residues glycine and proline are mapped on alternative Ramachandran plots because of their alternative flexibility characteristics. Ramachandran plots are frequently used as a visual method of structural validation.

Chapter 3 – Characterisation of substrate-free neprilysin

Foreword

This chapter describes the expression, purification and crystallisation of the neprilysin substrate-free crystal structure. The resulting substrate-free crystal structure is included as a published manuscript - *Moss S, Subramanian V & Acharya KR (2018) High resolution crystal structure of substrate-free human neprilysin. J. Struct. Biol. 204, 19–25*. More extensive background work has been described as a commentary text in addition to a more detailed introduction with the aim of providing greater context to the work within this thesis. A comprehensive methods section is also included in the commentary text for completeness.

3.1 Introduction

3.1.1 Neprilysin

Neprilysin (NEP, EC 3.4.24.11) is a transmembrane zinc metalloprotease composed of 749 amino acids spread across three domains. Short intracellular and transmembrane domains contain 27 and 23 amino acids respectively, while the large catalytic extracellular domain contains 699 amino acids (Figure 3.1).

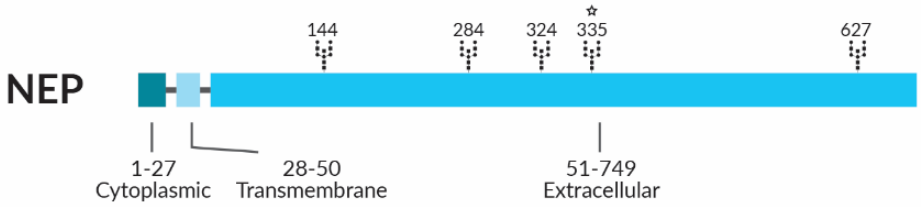


Figure 3.1. NEP domain organisation and known glycosylation sites. Glycosylation site 335 is predicted but has yet to be supported by experimental evidence.

To date, eight inhibitor-bound NEP extracellular domain crystal structures have been reported (Oefner *et al.,* 2000; Oefner *et al.,* 2004; Sahli *et al.,* 2005; Oefner *et al.,* 2007; Glossop *et al.,* 2011; Schiering *et al.,* 2016). These structures have revealed the domain is ellipsoid in shape and composed largely of α-helices (Figure 3.2a). Structural data has also uncovered the presence of four surface N-linked glycosylation sites; the importance of

which has been investigated through mutagenesis experiments. Removal of these four sites completely eliminates NEP activity, however activity can be restored close to maximal with glycosylation at either N144 or N627 (Lafrance *et al.*, 1994; Sato *et al.*, 2012). Detailed structural studies have further revealed the presence of a large enclosed cavity located at the centre of NEP. It is inside this cavity that the catalytic site is located, and substrate cleavage takes place.

The catalytic site of NEP is based around the conserved HXXEH zinc binding motif (Cerdà-Costa and Gomis-Rüth, 2014) (Figure 3.2b). Here a zinc ion is coordinated by conserved residues H583 and H587 in addition to residue E646. The conserved E584 residue is positioned above the zinc ion and is responsible for the generation of hydroxide ions in the proposed catalytic mechanism.

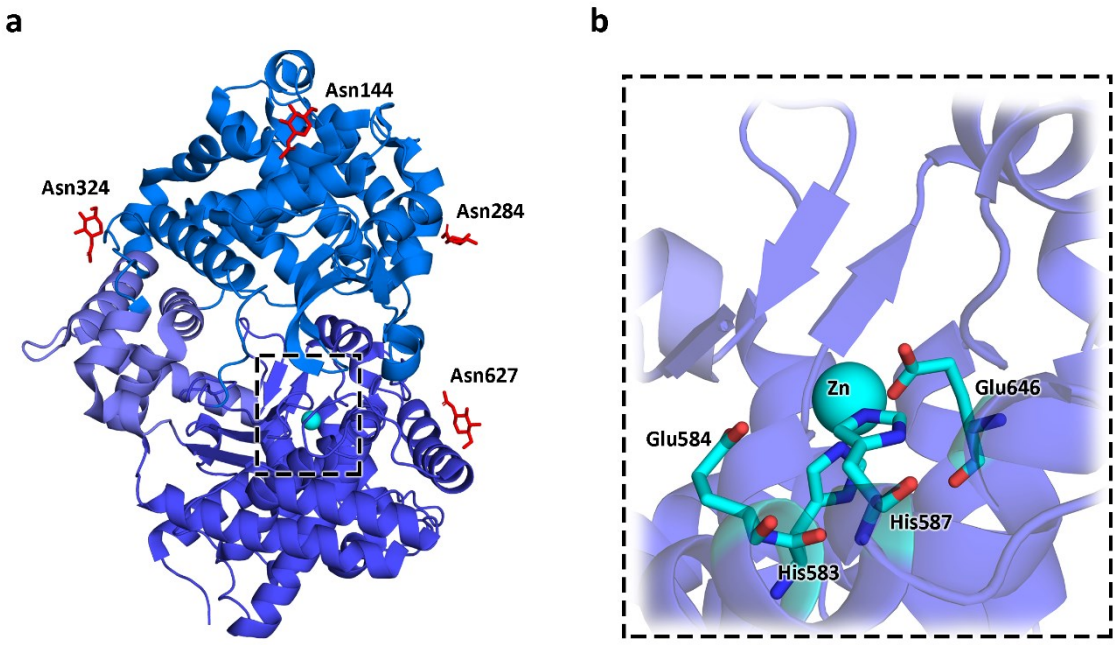


Figure 3.2, The extracellular domain of NEP. (a) cartoon illustration of NEP with four confirmed glycosylation sites indicated in red. (b) the catalytic site of NEP located around a zinc ion shown in cyan. Here, conserved residues H583 and H587, in addition to E646, coordinate the zinc ion. Conserved E584 is involved in the proposed catalytic mechanism.

The eight published NEP inhibitor-bound structures have been used to characterise inhibitor binding. Inhibitors of NEP are peptide-like and display a similar mode of binding at the catalytic site. That is, inhibitors coordinate the catalytic zinc ion and display similar

interactions to proximal NEP residues. The peptide-like nature of NEP inhibitors can also provide information on likely interactions involved in substrate binding. While much information has been gained from these structures, the dynamics involved in binding remain unclear. Does binding induce any domain or specific conformational changes, or does binding have a requirement for active site flexibility? These questions make determination of a substrate-free structure desirable. A substrate-free structure would provide information on the protein in an unbound state which, through comparison with the bound form, could reveal any movements that occur as a result of substrate or inhibitor binding.

There is a second basis for which a substrate-free structure is desirable. To date, all structures of inhibitor-bound NEP have been determined in a closed conformation. In this form, the central cavity is occluded and inaccessible to substrates. It is unclear whether the observed closed conformation is induced by inhibitor binding, a symptom of crystal packing or the natural conformation that exists in solution. With a substrate-free structure it would be possible to identify if ligand-binding was responsible for the observed closed protein conformation.

3.1.2 Protein expression in Pichia pastoris

Structural investigation using X-ray crystallography requires large amounts of protein (>1 mg) in order to screen conditions for protein crystallisation. Four major expression systems are available for production of recombinant proteins (Palomares *et al.,* 2004). These are bacterial, yeast, insect and mammalian. It is common practice to first explore expression in bacterial systems due to the fast culture growth, simple culture conditions, low cost and high quantity of protein production. Despite the advantage's, bacterial systems are frequently inappropriate for protein expression particularly those of eukaryotic origin. This is largely due to solubility issues or the absence of correct post-translational modifications.

NEP is a glycosylated protein that has been shown to require glycosylation for activity. It is most likely that glycosylation is required during protein folding and the absence of which does not produce a biologically relevant protein suitable for experimental studies. Because of this, expression of NEP in a bacterial system is inappropriate and previously published crystal structures of NEP have generated the protein in yeast or insect expression systems (Oefner *et al.,* 2000; Schiering *et al.,* 2016).

Pichia pastoris is a yeast expression system commonly utilised to produce glycosylated mammalian proteins. The expression system is well documented and commercially available from Invitrogen (EasySelect *Pichia* Expression Kit). Protein expression in *P. pastoris* is achieved by integrating the gene of interest into the genome under control of the alcohol oxidase (AOX1) promoter (Ellis *et al.,* 1985; Cregg *et al.,* 1993). Protein expression is then induced when methanol is added to the growth media. Proteins expressed in *P. pastoris* benefit from a higher eukaryotic protein processing, protein folding and post-translational modifications. Another main benefit of protein production in *P. pastoris* is the produced protein can be secreted. The low levels of natively secreted proteins can assist in the ease of purification.

P. pastoris has a large advantage over other yeast systems due to the lack of hyperglycosylation. Common baker's yeast, *Saccharomyces cerevisiae*, adds 50-150 mannose residues to an appropriate asparagine site (Asn-X-Ser/Thr) compared to the 8-14 residues observed in *P. pastoris* (Tschopp *et al.,* 1987; Grinna and Tschopp, 1989). The reduced level of glycosylation is beneficial when attempting to crystallise a protein of interest. In general, the presence of large flexible sugar chains on the outside of a protein may greatly reduce the propensity for crystallisation through reduction in the number of close packing contacts in the crystal lattice.

3.2 Methods

All reagents for expression and purification were sourced from Sigma-Aldrich or Fisher Scientific unless otherwise specified.

3.2.1 General experimental methods

All buffers were made in-house unless otherwise stated. The pH of buffers was determined using a Jenway 3510 pH meter and a Jenway pH temperature electrode. The pH for all buffers was measured at 18°C. Calibration of the pH meter was performed using buffer standards at pH 4, 7 and 10 sourced from Fisher Scientific (product numbers 10427260, 10000642, 10774074).

Aseptic technique was used when manipulating any microorganism to avoid contamination. This work was performed next to an open flame using sterile plasticware, glassware and cell growth media. Sterile plasticware including falcon tubes, petri-dishes and cell spreaders were purchased. Glassware and cell growth media were autoclaved in-house unless stated otherwise. Sterilisation by autoclave was performed at a minimum temperature of 121°C for 30 minutes at 15 psi. Filter sterilisation of solutions was performed by filtering through a 0.22 µm filter membrane (Millipore) into a sterile container next to an open flame.

3.2.2 Buffers

***P. pastoris* growth media**	
Yeast Extract Peptone agar (YPD agar)	1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose and 2% (w/v) agar
1 M potassium phosphate buffer, pH 6.0	132 ml of 1 M K_2HPO_4, 868 ml of 1 M KH_2PO_4 (pH 6.0, adjusted with phosphoric acid or KOH)
Buffered complex medium containing glycerol (BMGY)	1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate buffer (pH 6.0), 1.34% (w/v) yeast nitrogen base, 4×10^{-5}% (w/v) biotin and 1% (v/v) glycerol. Yeast nitrogen base, biotin and glycerol were filter sterilised and added to the other components which had been autoclaved.
Buffered complex medium containing methanol (BMMY)	1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate buffer (pH 6.0), 1.34% (w/v) yeast nitrogen base, 4×10^{-5}% (w/v) biotin and 0.5% (v/v) methanol. Yeast nitrogen base, biotin and methanol were

	filter sterilised and added to the other components which had been autoclaved.
SDS-PAGE buffers	
20x Tris-MOPS running buffer	1.2 M Trizma base, 0.6 M MOPS, 2.0% (w/v) sodium dodecyl sulfate.
4x SDS-PAGE sample buffer	250mM Trizma base (pH 6.8, pH adjusted with HCl), 10% (w/v) sodium dodecyl sulfate, 0.008 % (w/v) bromophenol blue, 0.588 M β-mercaptoethanol, 40% (v/v)glycerol.
Purification buffers	
Binding buffer	25 mM Trizma base, 150 mM NaCl, 2 mM $MgCl_2$ (pH 7.5) (pH adjusted with HCl).
Elution buffer	25 mM Trizma base, 150 mM NaCl, 2 mM $MgCl_2$, 500 mM imidazole (pH 7.5) (pH adjusted with HCl).
Crystallisation buffer	
JCSG-plus A12 (Molecular Dimensions)	0.2 M potassium nitrate and 20% (w/v) PEG3350

Table 3.1, Growth media and buffers required for the expression, purification and crystallisation of NEP.

3.2.3 NEP P. pastoris clone

A clone of *P. pastoris* GS115 containing a gene encoding the NEP extracellular domain (residues Tyr51-Trp749) was purchased from Invitrogen. The gene was integrated into the genome under control of the AOX1 promotor. The integrated gene also encoded an N-terminal 6x histidine tag and signal secretion sequence that would be cleaved after protein secretion had occurred.

3.2.4 Protein expression

Expression of the *P. pastoris* clone was done in accordance with the Invitrogen EasySelect expression manual (cat. no. K1740-01). A YPD agar plate was inoculated with the glycerol stock. The plate was incubated for 48 hours at 30 ºC to establish colonies of the *P. pastoris*. A single colony was used to inoculate a starter culture containing 25 mL of BMGY media in a 250 mL baffled flask. The starter culture was incubated overnight at 30 ºC 225rpm. 10 mL

of the overnight culture was used to inoculate a 250 mL BMGY media in a 2 L baffled flask. The main culture was incubated at 30 ºC and 225rpm for 24 hours.

After 24 hours the main culture was transferred to 500 mL centrifuge tubes which were spun at 6000 RCF for 20 minutes at 19 ºC. The supernatant was discarded, and the cell pellet was resuspended in 250 mL of BMMY media for induction. Resuspended cells were transferred to a new 250 mL baffled flask. The culture was then incubated for 72 hours at 30 ºC 225rpm. Every 24-hours 2.5 mL of 50% (v/v) methanol was added to each 250 mL culture.

After 72 hours of induction the main culture was transferred to a 500 mL centrifuge tubes and spun at 6000 RCF for 20 minutes at 19 ºC. The supernatant was collected, and the cell pellet was discarded. Solid Trizma base and NaCl were added to the supernatant to give final concentrations of 25mM Trizma and 150mM NaCl. The supernatant was mixed with magnetic stirrer for 10 minutes at room temperature. After this time the supernatant showed some precipitation which was removed by flowing through a 0.22 µm filter (Millipore). The final filter supernatant had a pH of ~7.

3.2.5 Protein purification

The buffer equilibrated *P. pastoris* supernatant was loaded onto a 5 mL HisTrap HP affinity column (GE Healthcare Life Sciences) at 2 mL/min pre-equilibrated with binding buffer. Following loading, the column was washed with binding buffer until the UV trace returned to baseline. A gradient elution (0 – 100 % over 30 minutes) of elution buffer at 5 mL/min was used to elute the bound protein. This method was later optimised to a stepwise elution. For this 2% elution buffer was used to wash the column until UV trace stabilised. Bound neprilysin was eluted using a single step of 50% elution buffer and 2 mL fractions were collected.

Protein containing fractions were confirmed using SDS-PAGE gel electrophoresis. These fractions were concentrated to under 2 mL in a centrifuge at 4000 RCF using a 30kDa molecule weight cut off (Millipore) filter unit. The concentrated protein containing sample was then loaded at 5 mL/min onto a HiPrep 26/10 Desalting column (GE Healthcare Life Sciences) pre-equilibrated with binding buffer. Alternatively, a Superdex 16/60 was used, pre-equilibrated with binding buffer at 1.5 mL/min. Protein containing fractions were stored at 4 ºC for short term storage or -20 ºC for longer term storage.

3.2.6 SDS-PAGE gel electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to determine purity of protein samples and approximate protein molecular weights. Poly acrylamide gels 'TruPAGE Precast Gels 10%' (Sigma-Aldrich) were used with 1x Tris-MOPS running buffer.

Protein samples were added to SDS-PAGE sample buffer (1x final concentration) and were electrophoresed with 'Pageruler Plus Prestained Protein Ladder' (ThermoFisher) or 'Protein Marker Broad Range 2-212 kDa' (New England BioLabs). Electrophoresis was conducted using a constant 150v for 45 minutes or until the dye front reached the bottom of the gel. Poly acrylamide gels were stained in SimplyBlue SafeStain (ThermoFisher) and destained in distilled H_2O.

3.2.7 Mass spectrometry

Purified NEP was buffer exchanged into water and made to 1 mg/mL for HPLC using a centrifuge filter unit with a 30kDa molecule weight cut off (Millipore). The Chip-based analysis was conducted using an HPLC-Chip Cube system coupled to a 6520 quadrupole time-of-flight (QTOF) mass spectrometer (Agilent Technologies, Santa Clara, CA) operated in ESI positive-ion mode. Liquid chromatography was performed using a Protein-Chip (II) with a 40 nL enrichment column and analytical column of 43 mm x 75 µm with Zorbax 300SB-C8 packing material at 5 µm (G4240-63001, Agilent, Santa Clara, CA). The ChipCube source was operated at 365 ºC with 5 L^{-1} N_2 drying gas, the capillary voltage set to 2100V and fragmentor at 400V. The source was interfaced with an Agilent 1260/1200 series HPLC system consisting of a 1260 Cap pump, 1200 Nano pump, 1200 Micro WPS and 1290 Infinity Thermostat (Agilent, Santa Clara, CA). Between 0.2 and 1 µL samples were loaded onto the enrichment column using the capillary pump flow with H_2O + 0.1 % (v/v) formic acid at a flow rate of 4 µL/min. The samples were eluted onto the analytical column using the nano pump at a flow rate of 0.6 µL/min. Solvent A and B consisted of H_2O + 0.1 % (v/v) formic acid and acetonitrile:H_2O 90:10 with 0.1 % (v/v) formic acid. Gradient steps were as follows: 0-4 min from 3 % B to 50 % B, 4-5 min to 100 % B, 5-11 min 100 % B, 11-12 min from 100 % to 3 % B. Internal lock mass calibration during the run was done using one calibration reference mass at 1221.9906 m/z. Data processing was performed using the Masshunter Workstation software version B.50.00 (Agilent, Santa Clara, CA, USA).

3.2.8 Deglycosylation

Deglycosylation of NEP was conducted using the Endoglycosidase H (Endo H, Cat. No. P0702S NEB). Deglycosylation experiments were conducted in accordance with the user manuals under either native or denaturing conditions.

Native: 20 µg purified NEP was combined with 2 µl Endo H (1000U) and 4 µL 10x Glycobuffer 3 (NEB) before making up to 40 µL with H_2O. The solution was incubated 37ºC for two hours before the sample was analysed using gel electrophoresis.

Denaturing: 20 µg purified NEP was incubated at 100ºC for 10 minutes in 1X Glycoprotein denaturing buffer (NEB). After cooling, the solution was combined with 2 µL Endo H (1000U) and 4 µL 10x Glycobuffer 3 (NEB) before making up to 40 µL with H_2O. The solution was incubated 37ºC for two hours before the sample was analysed using gel electrophoresis.

3.2.9 Cleavage assay

Fluorogenic substrate Mca-RPPGFSAFK-(Dnp) (Enzo Life Sciences) was used to confirm the expressed NEP protein displays protease activity. The 7-methoxycoumarin (Mca) group is efficiently quenched by resonance energy transfer to the 2,4-dinitrophenyl (Dnp) group. Peptide cleavage of an amide bond between the two groups allows fluorescence of the Mca. Fluorescence was assessed using a CLARIOstar high-performance microplate reader (BMG LABTECK). Excitation of the Mca group was achieved at 328 nm and emission was detected at 393nm and readings were taken at 15 second intervals for varied total durations.

Mca-RPPGFSAFK(Dnp) was dissolved in DMSO to 1 mM and stored at -20 ºC. For the cleavage assay, 50 µL of purified neprilysin (varied concentrations) was added to a black 96 well microplates (Greiner Bio-One). Mca-RPPGFSAFK(Dnp) DMSO stock was diluted to 20 µM in purification binding buffer for NEP. 50 µL of Mca-RPPGFSAFK(Dnp) was added to the well, to give a final concentration of 10 µM using the CLARIOstar inbuilt injection needle. This provided mixing of the components.

Assays were repeated three times where sample availability permitted. Assay data is shown after subtraction of baseline fluorescence using 10 µM Mca-RPPGFSAFK(Dnp). Assay figures were produced in Graph Pad Prism and standard deviation bars are shown.

3.2.10 Crystallisation

Purified NEP was concentrated by centrifugation at 4000 RCF to 4 - 12 mg/mL using centrifuge filter unit with a 30kDa molecule weight cut off (Millipore). Concentrations were determined using a NanoDrop (Thermo Scientific) and the extinction coefficient of 1.57 ($A^{0.1\%}_{280}$ value) for NEP (calculated using ProtParam). A pre-crystallisation test (Hampton research, Cat. No. HR2-140) was used to confirm a concentration of 7.5 - 12 mg/mL was appropriate for crystallisation screening.

Initially crystallisation conditions were screened using a sitting-drop vapour diffusion set up with 96-well screens (Molecular Dimensions). Sitting drop crystallisation screens were performed using a robot (Phenix, Art Robbins Instruments, USA), 96-well plates (Intelli-plate, Art Robbins Instruments) and ClearVue plate seals (Molecular Dimensions). 50 µL of crystallisation screen was used for the reservoir solution. The sitting drop was formed from 0.18 µL of protein and 0.18 µL of reservoir solution. Plates were incubated at 18 ºC and checked weekly for crystal formation.

24-well hanging drop crystallisation screens were used to optimise conditions that showed crystal formation in the 96-well screens. Hanging drop screens were performed with 24-well plates (Greiner Bio-One) with glass cover slide (Greiner Bio-One) and high-performance vacuum grease (Dow Corning). 500 µL of screen condition was used as the reservoir solution. The hanging drop was formed from 1 µL of protein and 1 µL of reservoir solution. Plates were incubated at 18 ºC and checked weekly for crystal formation.

3.2.11 Synchrotron data-collection and processing

Protein crystals were harvested using micro loops of 20 µm to 400 µm in diameter (MiTeGen). Looped crystals were flash-cooled in liquid nitrogen and transferred to Diamond Light Source (Didcot, UK) for the collection of X-ray diffraction data. Beamlines i03 and i04 were used to collect the X-ray diffraction data and were fitted with PILATUS-6M detectors at the time of collection.

Strategies for data collection were determined using three test shots at 45º degree increments and software package EDNA (Incardona *et al.,* 2009). Between 3600 - 9999 images with 0.1º of oscillation were collected for a full data collection. The raw X-ray diffraction data were indexed and integrated using DIALS (Waterman *et al.,* 2016). Data were

scaled and merged using AIMLESS within the CCP4 suite (Evans *et al.,* 2013;Winn *et al.,* 2011). During scaling, bad images were removed and a resolution cut off was applied based on statistics generated from AIMLESS (Evans *et al.,* 2013).

Edge scans were conducted on i04 by exploiting the variable wavelength. Specifically a zinc metal scan was conducted between 9559-9749 eV and CHOOCH (Evans and Pettifer, 2001) was used to plot the anomalous-scattering factors f'' and f'.

3.2.12 Structure model building and refinement

Phaser (McCoy *et al.,* 2007) was used to perform molecular replacement using structure 5JMY (Schiering *et al*. 2016) on the merged and scaled data from AIMLESS. Structural solutions from Phaser were assessed using the LLG score and visual inspection of electron density maps. After Phaser, the structure factors and initial coordinates were refined using a combination of REFMAC5 (Murshudov *et al.,* 2011) and Phenix refine (Afonine *et al.,* 2012) from the Phenix suite (Adams *et al.,* 2010). Visualisation and alteration of the protein structure was done using Coot (Emsley *et al.,* 2010). Refinement of the protein structures was performed until successive round of refinement failed to improve R-factor statistics. Programmes MolProbity (Chen *et al.,* 2010) and PDB validation (Berman *et al.,* 2003) were used to assess any stearic and geometry outliers within the modelled structure. Figures were created using either PyMOL (Schrödinger, 2015) or CCP4mg (McNicholas *et al.,* 2011).

3.4 Results

3.4.1 Expression and purification

A clone of *P. pastoris* GS115 containing a gene encoding the extracellular domain of NEP (residues Tyr51-Trp749) was purchased from Invitrogen. The gene had been integrated into the genome under control of promoter AOX1 and contained an additional N-terminal histidine tag and secretion signal. A protocol for recombinant protein expression in *P. pastoris* is included in the Invitrogen EasySelect expression manual (Cat. No. K1740-01). Briefly, expression begins with a 25 mL overnight culture inoculated from an agar plate. The overnight culture is used to inoculate a 1 L expression culture in growth media (BMGY). After 16-24 hours the expression culture is centrifuged and resuspended in the same volume of induction media containing 0.5% (v/v) methanol (BMMY). After every 24-hours 2.5 mL of 50% (v/v) methanol is added to maintain the 0.5% (v/v) concentration until the supernatant is harvested.

The described expression protocol provides a guideline for protein expression but through parameter modification protein production can be increased. Target parameters included cell density prior to induction, volume of induction media and methanol concentration. These parameters were investigated using smaller 250 mL expression cultures to optimise protein expression before beginning large scale expressions. The varied expression conditions are shown in Table 3.2. The expressed NEP contained a secretion tag, as such, quantity of protein expressed was determine by sampling the supernatant before and after induction and analysing samples with SDS-PAGE gel electrophoresis (Figure 3.3).

Pre-induction supernatant samples contained no observable protein bands on an SDS-PAGE gel. The post-induction supernatant samples all contained a band present at ~100 kDa. While the calculated mass of the NEP extracellular domain is 81.8 kDa, the presence of glycosylation was expected to increase the observed molecular weight. As four N-linked glycosylation sites for NEP are known, and *P. pastoris* commonly adds 8-14 sugars per site, the total mass was expected to be 6 – 10 kDa larger than just the protein mass alone. The additional mass and the absence of protein bands in pre-induction samples strongly suggested the band at ~100 kDa corresponded to expressed NEP.

While all expression conditions contained this band corresponding to NEP, expression levels varied. Equally intense bands were observed in the conditions 1 and 3. Condition 2 had a less intense NEP band and several other bands were present that may correspond to degradation products. Neither changes to induction volume in conditions 4 nor 5 improved protein expression over standard conditions. Based on this, condition 1 was selected for future protein expressions.

Condition	**1**	**2**	**3**	**4**	**5**
Inoculation volume (mL)	10	**40**	10	10	10
Cell growth media (BMGY) volume (mL)	250	250	250	250	250
Cell induction media (BMMY) volume (mL)	250	250	250	**500**	**125**
Methanol concentration (%, v/v)	0.5	0.5	**1.0**	0.5	0.5

Table 3.2, NEP protein expression optimisation in P. pastoris. Five different expression conditions were screened to determine if variation in cell density prior to induction, volume of induction media and methanol concentration resulted in greater protein production.

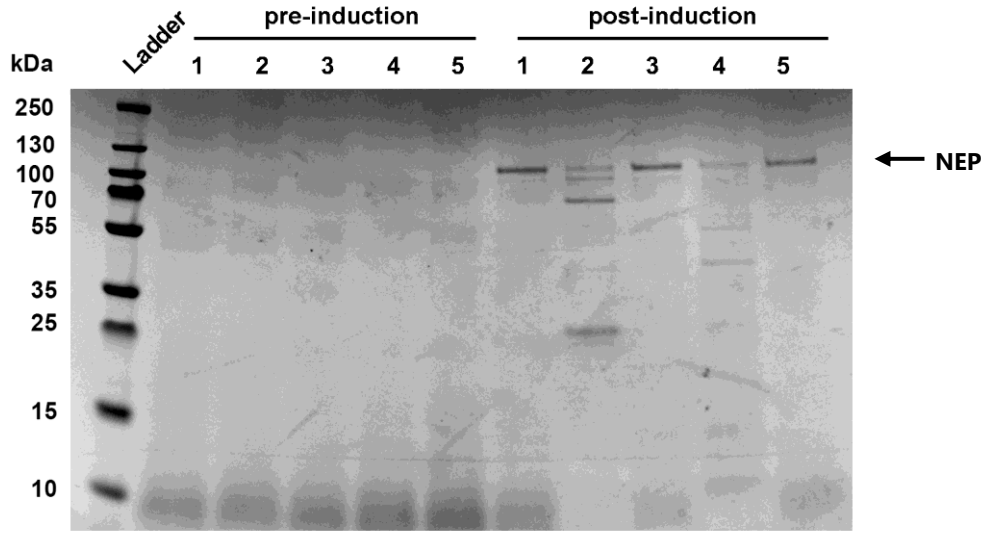


Figure 3.3, NEP expression optimisation in P. pastoris. Supernatant samples taken from expression cultures before and after induction are analysed using SDS-PAGE gel electrophoresis. A band corresponding to NEP is present at ~100 kDa and is observed at varying levels in all conditions post induction.

NEP expression was repeated on a larger scale and a suitable purification method was developed. Given that NEP was secreted from *P. pastoris* and contained a 6x histidine tag

the first purification step was performed by loading the harvested supernatant onto a histidine trap affinity column (GE Healthcare). The bound protein was eluted using a gradient of imidazole containing elution buffer. A single protein peak corresponding to NEP and was eluted at approximately 15% elution buffer. This gradient elution was later switched to stepwise elution to reduce purification time and elution volume. The stepwise elution contained a wash step of 2% elution buffer followed by a protein elution step of 50% elution buffer (Figure 3.4).

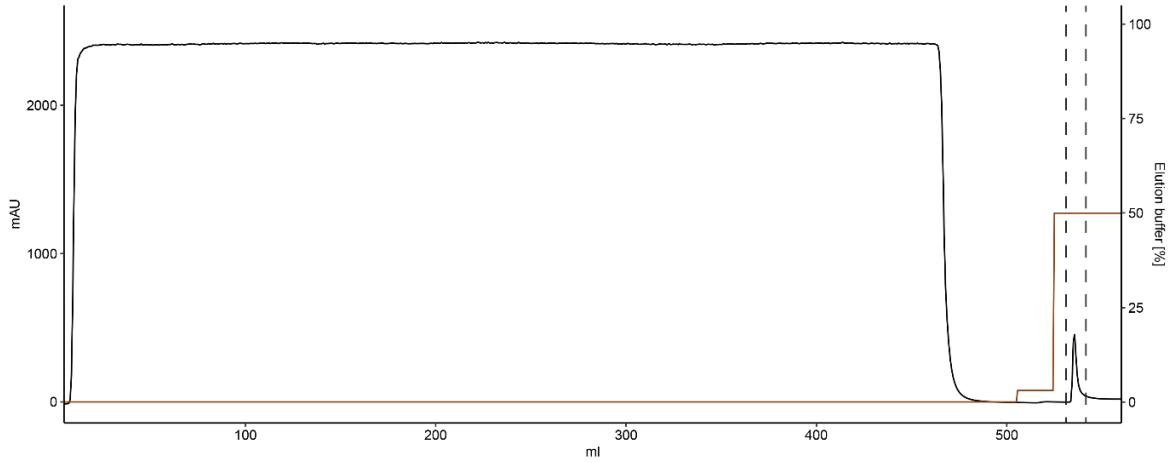


Figure 3.4, NEP histidine affinity trap purification. The protein containing supernatant from a P. pastoris culture was loaded onto a histidine affinity trap column (GE healthcare). Once loaded the column was washed with binding buffer followed by a wash step of 2% elution buffer. Protein elution was achieved at 50% elution buffer. A single peak was observed and peak containing fractions (indicated between dash lines) were pooled.

Peak containing fractions from the histidine affinity column were observed to be light brown in colour, similar to that of expression growth media. While SDS-PAGE gel electrophoresis revealed the protein was pure after the histidine affinity column purification step (Figure 3.7), discolouration in the fractions indicated unknown components were present. As the purified protein was intended for crystallisation trials, knowledge of the exact composition of the protein buffer was required to ensure the repeatability of future crystallisation hits. A second purification step was therefore performed using a HiPrep 26/10 desalting column (GE Healthcare) in order to buffer exchange the sample (Figure 3.5). This second purification step was later changed to a size exclusion purification step with a Superdex 16/60 column

(GE Healthcare) to improve purity (Figure 3.6). Both secondary purification steps were successful in removing the light brown colour from protein samples.

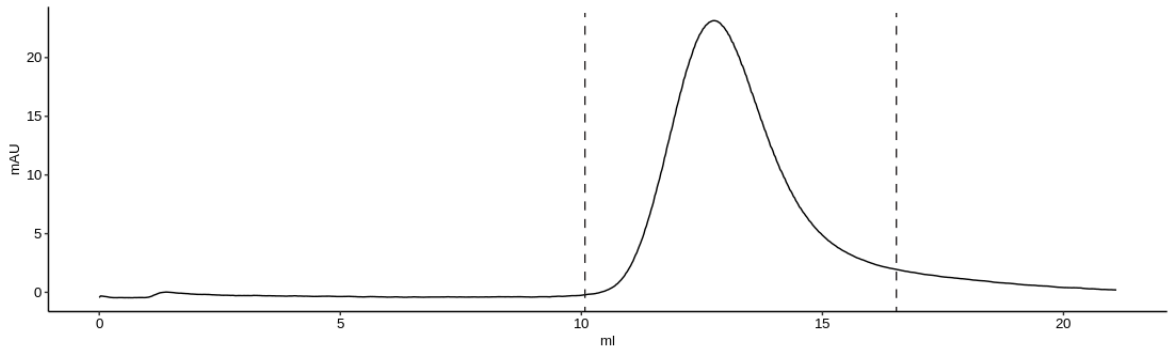


Figure 3.5, NEP buffer exchange purification. The pooled fractions eluted from the histidine affinity trap column were loaded onto a HiPrep 26/10 desalting column to remove impurities from the induction media and to buffer exchange the protein into binding buffer. A single peak was observed and peak containing fractions (indicated between dash lines) were pooled.

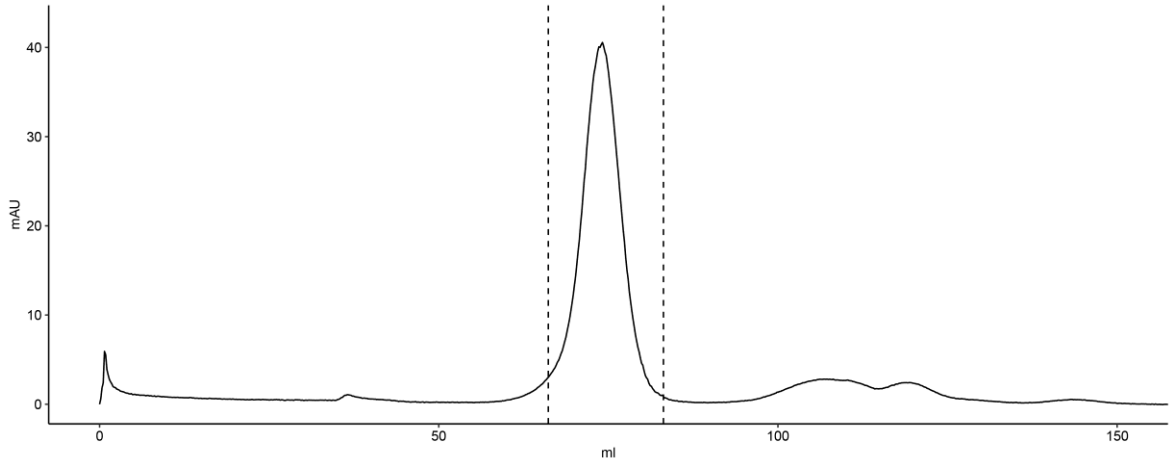


Figure 3.6, NEP size exclusion purification. The pooled fractions eluted from the histidine affinity trap column were loaded onto a Superdex 16/60 size exclusion column to remove impurities from the induction media and to buffer exchange the protein into binding buffer. A single peak was observed and peak containing fractions (indicated between dash lines) were pooled.

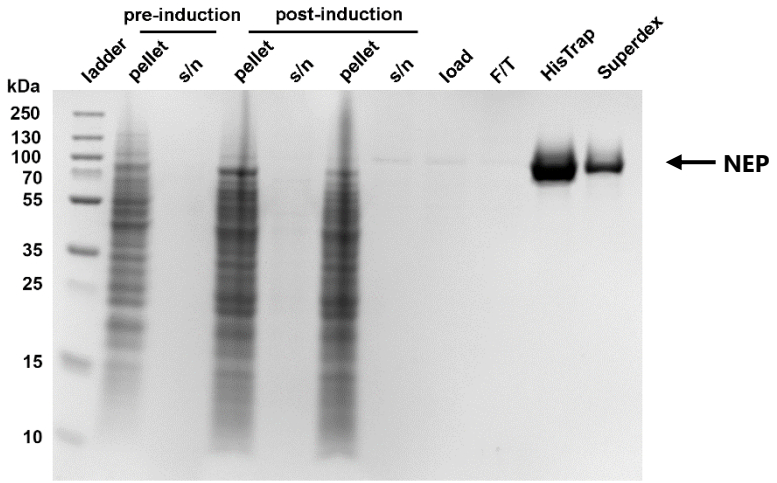


Figure 3.7, NEP expression and purification in P. pastoris. SDS-PAGE gel electrophoresis was used to analyse samples taken during NEP expression. Supernatant (s/n) and cell pellet (pellet) samples were taken from the P. pastoris expression culture at time points: 24-hours pre-induction, 24-hours post-induction and 72-hours post induction. Samples of the harvested supernatant (load) and HisTrap column flow through (F/T) are also shown. The eluted and pooled fractions from the HisTrap and size exclusion purification steps reveal the high purity of the final product.

Samples of the protein expression and purification were analysed SDS-PAGE gel electrophoresis to assess protein purity (Figure 3.7). It was clear from this analysis that high protein purity had been achieved in the final purified sample. The final purified protein was pooled, and the absorbance was recorded to indicate ~1 mg of protein could be purified from 1 L of *P. pastoris* culture.

3.4.2 Liquid chromatography-mass spectrometry

Liquid chromatography-mass spectrometry (LC-MS) was used to support SDS-PAGE analysis and confirmed that NEP had been successfully expressed and purified. As previously stated, the calculated mass of the NEP extracellular domain is 81.8 kDa however the presence four N-linked glycosylation sites result in a higher expected mass. This expected mass can be predicted as between 87.8 – 91.8 kDa given the knowledge that *P. pastoris* commonly adds 8-14 sugars per site. LC-MS analysis identified a range of species between 82kDa to 97kDa (Figure 3.8a). Figure 3.8b shows an enlarged section of the MS spectrum where individual peaks are separated by a mass equal to a single monosaccharide (~164 Da). The observed peaks strongly suggest that the purified protein is NEP which has been expressed in a range of glycosylation states. The majority of glycosylation states have a mass of between 89-

94kDa which corresponds to an average estimated glycosylation of 42-73 monosaccharide units per protein.

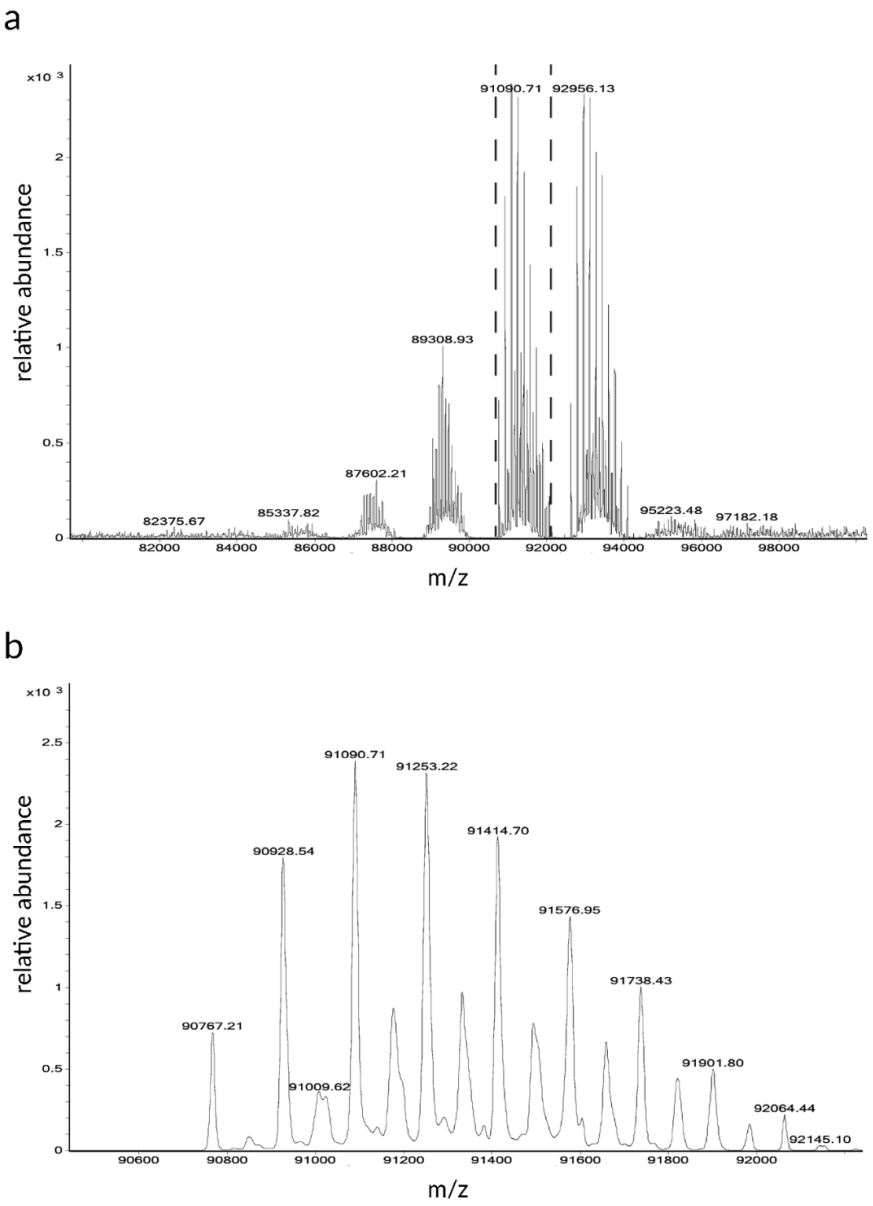


Figure 3.8, NEP mass spectrometry analysis. (a) deconvoluted mass spectrum of the protein glycosylation variants with different masses between 82 kDa to 97kDa. (b) enlarged section of the total mass spectrum. Peaks are separated by ~164 Da or a single sugar residue.

3.4.3 Cleavage assay

To confirm that the expressed and purified NEP protein was in an active and biologically relevant form the ability of the protein to cleave a fluorescent peptide was assessed. The chosen fluorescence peptide, Mca-RPPGFSAFK-(Dnp) (Enzo Life Sciences), contains a

fluorescent 'Mca' group that is quenched by the 'Dnp' group while in close proximity. Upon substrate cleavage release of the 'Mca' group allows recordable fluorescence to occur.

An initial assay was used to confirm peptide cleavage and determine an optimal concentration range that NEP could be used for accurate determination of initial rates. This range was determined to be between 1.6-8 nM (0.131-0.655 mg in the 100 µL reaction volume) (Figure 3.9). More extensive characterisation of kinetic parameters (K_m and V_{max}) are presented in Chapter 5. The peptide cleavage provided strong evidence that NEP, recombinantly expressed in *P. pastoris* and in the absence of transmembrane and cytoplasmic domains, was appropriate for crystallisation and structural analysis.

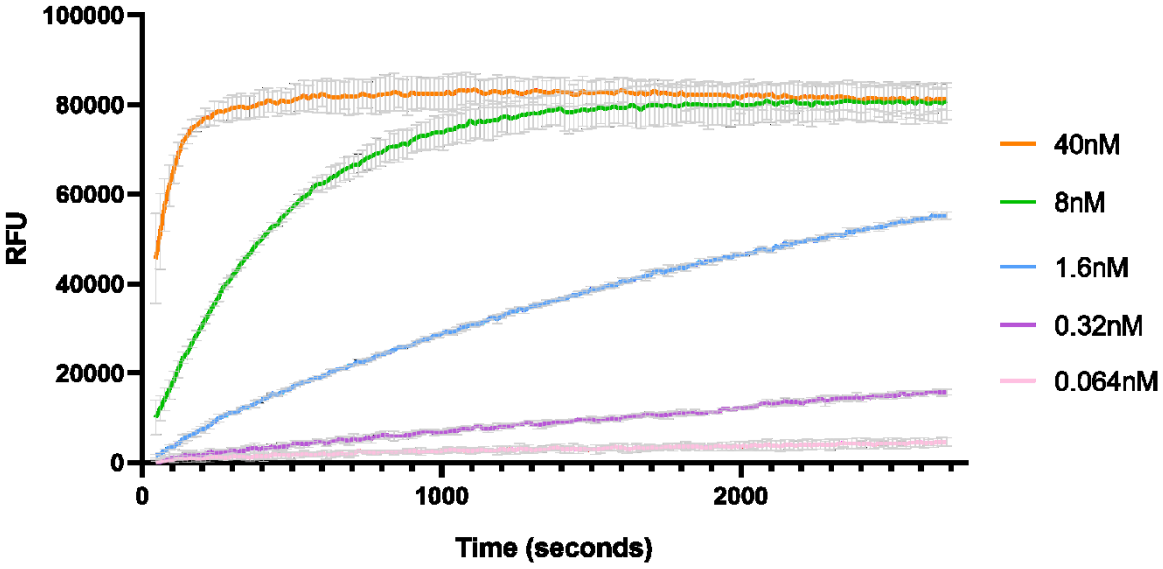


Figure 3.9, NEP peptide cleavage assay. Varied concentrations of NEP were incubated with 10 µM peptide substrate Mca-RPPGFSAFK-(Dnp). Curves are plotted with the mean value and ±SD error bars for three repeats.

3.4.4 Deglycosylation

The glycosylation of proteins frequently limits the generation of high-quality protein crystals. Increased surface flexibility, reduction in favourable crystal stacking interactions and the generation of heterogenous protein molecules can all be attributed to glycosylation and can be a detriment to crystallisation (Davis *et al.,* 2011). For these reasons all published structures of NEP to date, with the exception of 5JMY (Schiering *et al.,* 2016), have been

generated from protein that was expressed in a glycosylated form but was subsequently deglycosylated prior to crystallisation trials.

Using a *P. pastoris* expression system frequently reduces the amount of protein glycosylation compared with similar yeast expression systems. LC-MS analysis revealed this was true of the expressed and purified NEP where between 42-73 monosaccharide units per protein. Despite the relatively low level of glycosylation, deglycosylation of NEP was investigated to determine its practicality for crystallisation. This investigation was conducted alongside crystallisation of the glycosylated form and would only be required if glycosylated NEP did not crystallise.

Common glycosylation patterns in *P. pastoris* are well documented (Grinna and Tschopp, 1989). The general structure begins with two N-acetyl glucosamine (NAG) residues before the chain branches and has a high mannose content. Endoglycosidase (Endo H) is a commercially available glycosidase that removes high mannose glycans by cleaving between the first two NAG residues (Figure 3.10). After incubation with Endo H a glycoprotein contains just a single NAG at each glycosylation site.

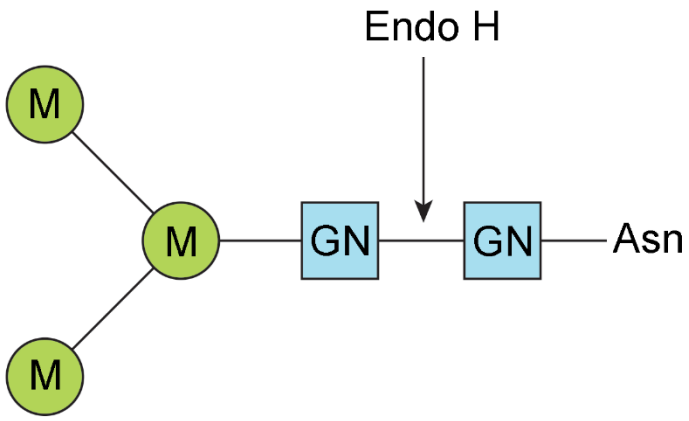


Figure 3.10, Protein N-linked glycosylation in P. pastoris. Two N-acetyl glucosamine (GN) sugars are bound to the asparagine residue. Attached to the N-acetyl glucosamine sugars are mannose sugars (M). Here a highly mannose region exists where length and branching can vary greatly. The cleavage site targeted by glycosidase Endo H is shown.

Deglycosylation experiments were achieved using the standard protocol and samples were analysed using SDS-PAGE gel electrophoresis to assess the removal of glycosylation (Figure

3.11). Deglycosylation with Endo H was successful and resulted in the observation of a protein band at a lower molecular weight. No difference was observed for Endo H treatment under protein denaturing and non-denaturing conditions suggesting that the same level of deglycosylation had occurred in both conditions.

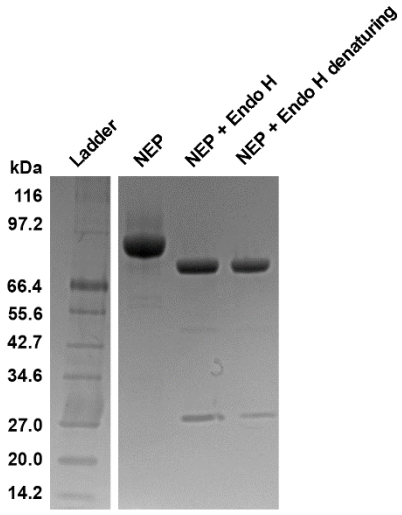


Figure 3.11, Deglycosylation of NEP. Endo H was incubated with NEP under native and denaturing conditions to assess deglycosylation. A decrease in molecular weight was caused with Endo H incubation but no difference was observed between the native and denaturing conditions.

Peptide cleavage of deglycosylated NEP was assessed to confirm the protein remained suitable for crystallography if required. Peptide cleavage was identified for deglycosylated NEP to a similar level to control protein for a single substrate concentration (Figure 3.12).

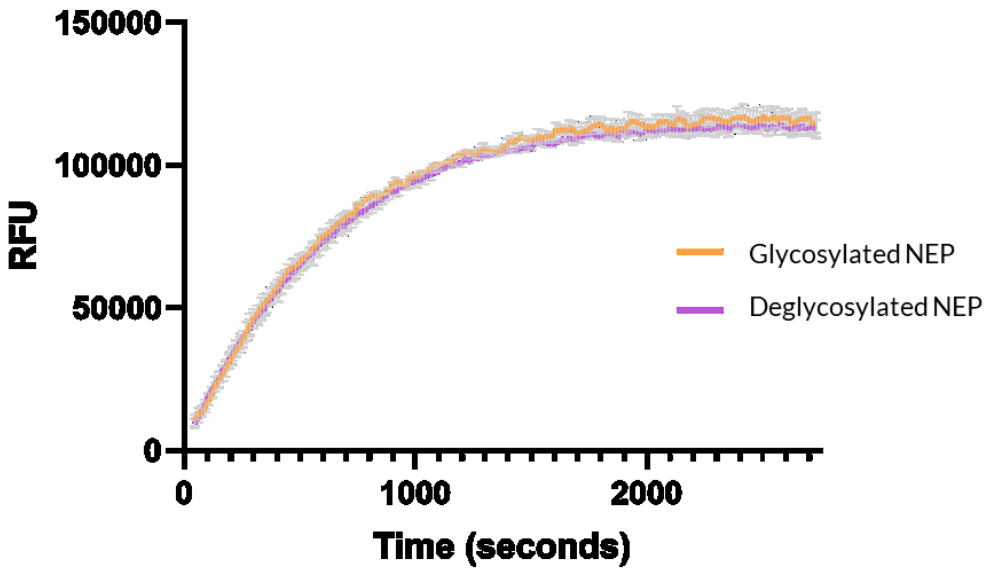


Figure 3.12, Deglycosylated NEP peptide cleavage assay. The cleavage of Mca-RPPGFSAFK-(Dnp) was assessed for glycosylated and endo H treated deglycosylated NEP using 0.655 mg of protein in the 100 µL reaction (8 nM). Both treated and untreated protein displayed vary similar assay profiles. Curves are plotted with the mean value and ±SD error bars for three repeats.

3.4.5 Crystallisation

Crystallisation of NEP had been reported previously using varied crystallisation conditions. As no crystal structure from a *P. pastoris* expression system was available, conditions for crystallisation were screened. Prior to high-throughput screening a Hampton research pre-crystallisation test was used to confirm 7.5 – 12.5 mg/mL as an appropriate concentration to begin crystallisation experiments.

Initially crystallisation trials were conducted on a glycosylated form of NEP. Nine high-throughput crystallisation trials were performed using 96-condition screens commercially available from Molecular Dimensions (Table 3.3). All screens were incubated at 18 ºC and were checked weekly.

Screen	**Crystallisation hits**
JCSG-plus	4
PACT premier	3
SG1	-
Structure Screen 1	-
Structure Screen 2	-
Morpheus	1
MemGold	1
Clear Strategy	-
MIDAS	-

Table 3.3, NEP crystallisation screens. A total of nine initial 96-well high throughput crystallisation screens were performed using the expressed and purified NEP. A total of 9 crystallisation hits were observed.

Crystallisation was observed with the condition JCSG-plus A12 (0.2 M KNO_3, 20% (v/v) PEG 3350) after a week. Other crystallisation hits were also identified (Figure 3.13), however only

condition JCSG-plus A12 was optimised as other crystallisation conditions did not form crystals for >4 months.

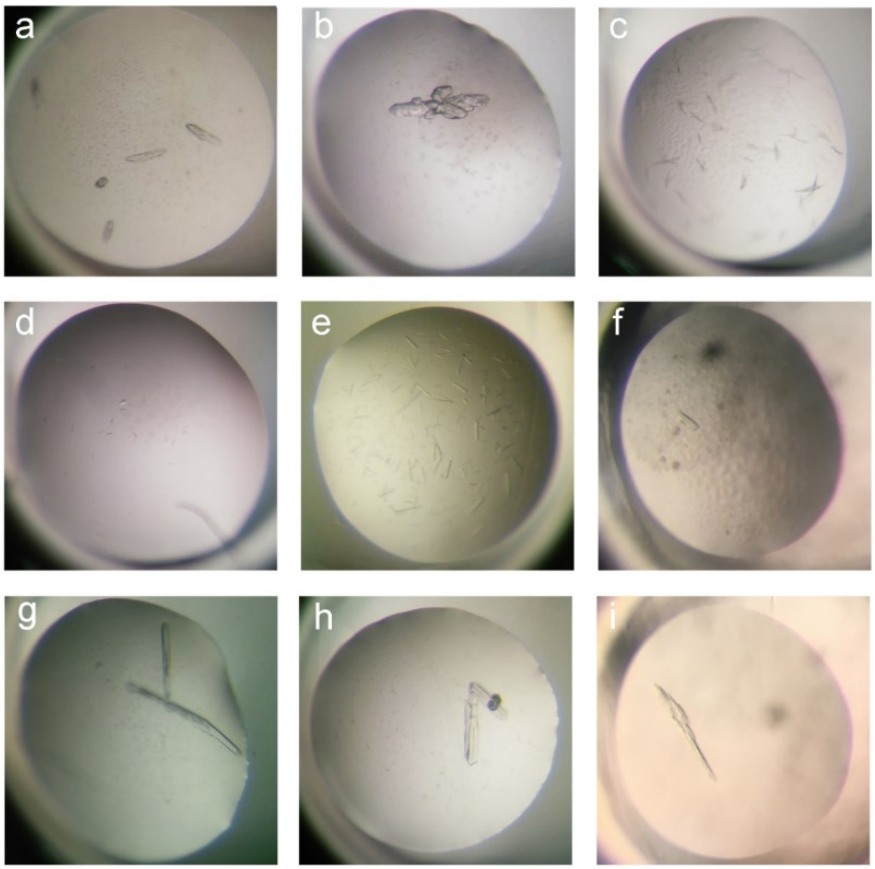


Figure 3.13, NEP crystallisation hits. Crystallisation hits were observed in conditions (a) JCSG-plus A12, (b) JCSG-plus C11, (c) JCSG-plus D10, (d) JCSG-plus E8, (e) MemGold B7, (f) PACT premier E4, (g) PACT premier E9, (h) PACT premier F5 and (i) Morpheus H2.

X-ray diffraction data was collected using crystals formed in the JCSG-plus A12. While these crystals diffracted, only low resolution (less than 4.5 Å) data could be collected (Figure 3.14). Despite this, it was possible to confirm the crystallised protein was NEP by solving the collected data using molecular replacement with an inhibitor-bound NEP structure (5JMY). It was also possible to confirm the presence of the catalytic zinc by exposing NEP crystals to a fluorescence scan at Diamond Light Source. The expected K-edge for zinc is 9658.6 eV which correlates well with the recorded K-edge (Figure 3.15)

Optimisation of condition JCSG-plus A12 was performed in a 24-well hanging drop set up with the aim of generating NEP crystals that diffracted to a higher resolution. Initially optimisation was attempted through variation in PEG 3350 concentration (15-25 % (v/v)). This optimisation was unsuccessful as higher concentrations formed large amounts of precipitate and reduced crystal formation, and lower concentrations failed to induce protein crystal formation. While optimisation of PEG 3350 concentration did not yield crystals of higher diffraction quality, the change from a 96-well crystallisation plate to a 24-well crystallisation frequently generated larger crystals which in turn diffracted to a higher resolution (Figure 3.16).

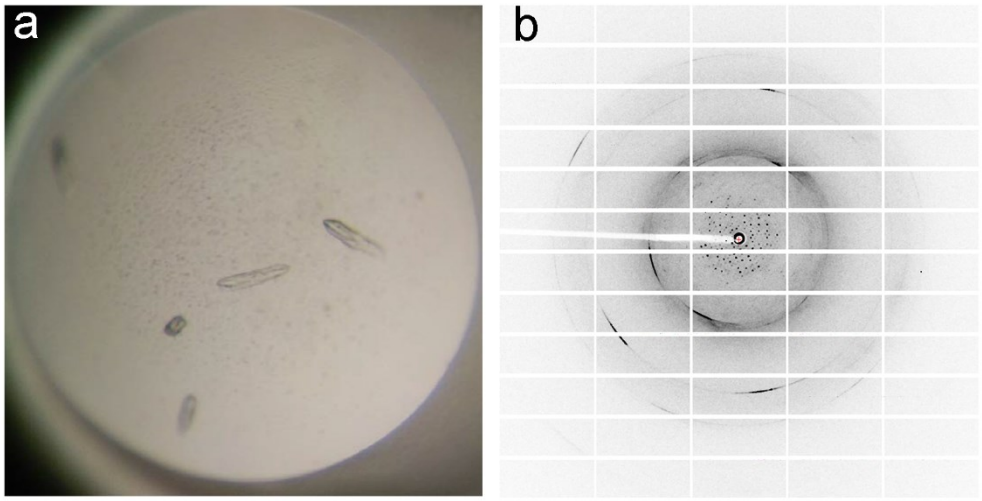


Figure 3.14, Initial crystallisation and diffraction of NEP. (a) NEP 96-well crystallisation hits from condition JCSG-plus A12 (0.2 M KNO_3, 20% (v/v) PEG 3350). (b) example screening diffraction image from NEP crystal (resolution is less than 4.5 Å).

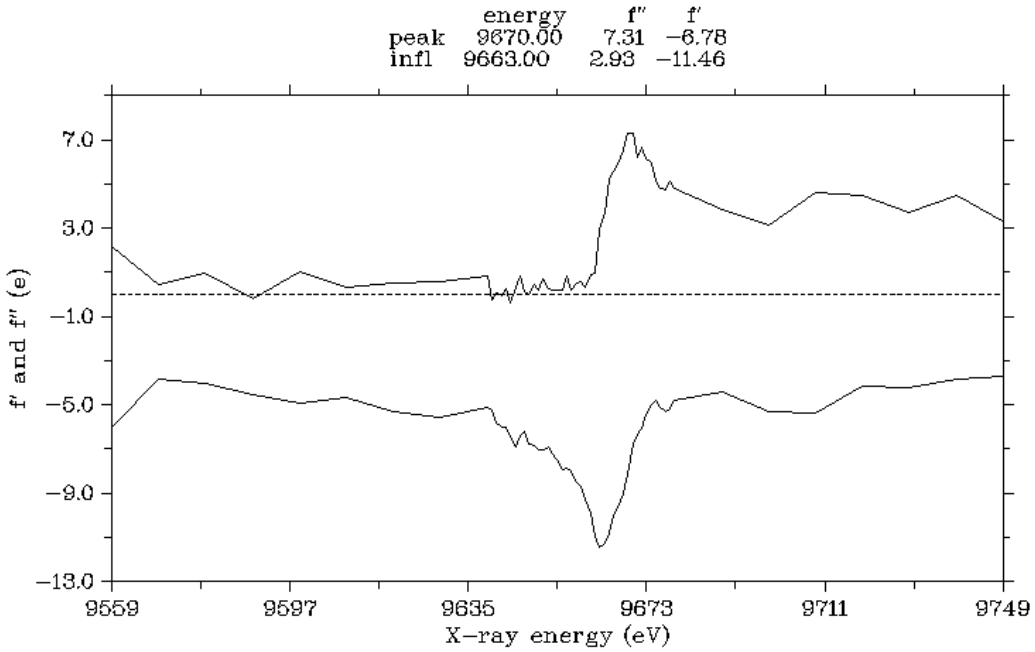


Figure 3.15, Fluorescence scan of NEP. CHOOCH plot shows the anomalous signal recorded at the zinc absorption edge for an NEP crystal. The observed signal confirms the presence of zinc within the NEP protein structure.

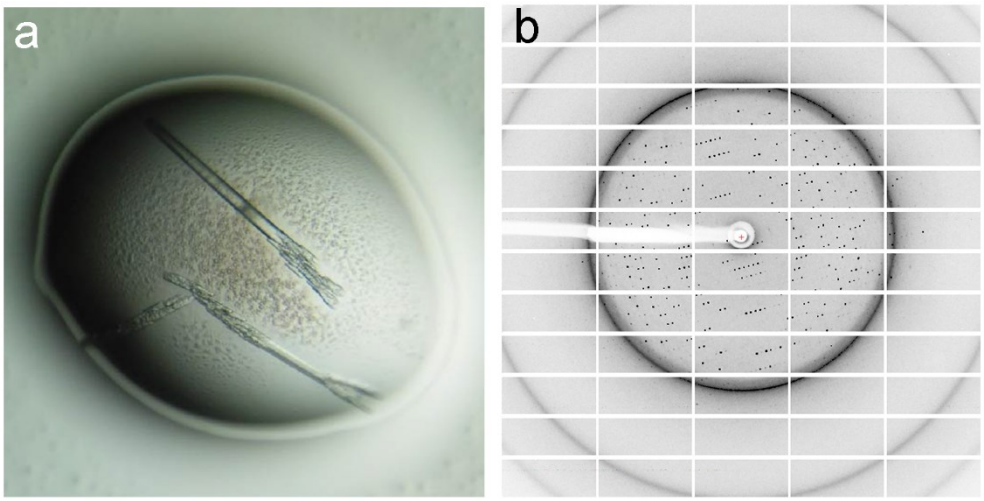


Figure 3.16, Optimised crystallisation and diffraction of NEP. (a) NEP 24-well crystallisation hits from condition JCSG-plus A12 (0.2 M KNO_3, 20% (v/v) PEG 3350). (b) screening diffraction image from NEP crystal (diffraction is higher than 2.5 Å).

Repeats of the 24-well crystallisation trials were performed with condition JCSG-plus A12 to generate NEP crystals. From these trials several observations were noted. On average 1:3

crystallisation drops formed crystals despite containing the same buffer and batch of purified protein. If crystallisation did occur diffraction quality varied a large amount between crystals. It was also clear that diffraction quality varied greatly between different areas of the same crystal (Figure 3.17). After screening over 100 crystals at Diamond Light Source less than 10 crystals diffracted to a resolution of 2.5 Å or better. The following published paper reports the highest-resolution structure for NEP that was obtained through crystal screening.

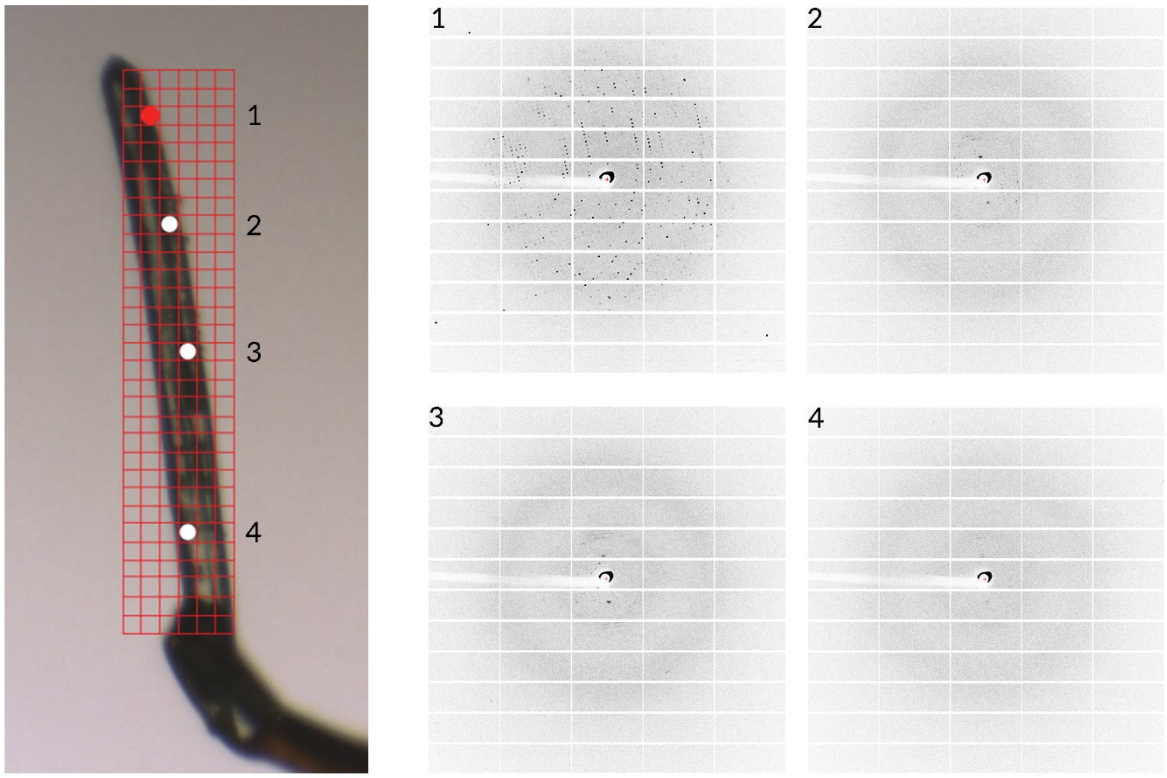


Figure 3.17, Diffraction variation across NEP crystals. Diffraction of NEP crystals varied dramatically across the length of a crystal. A grid scan was used to screen small areas of and NEP crystal. The example diffraction images show only one area of a crystal diffracted to a high quality.

Published article

Elsevier copyright policy:

"As an Elsevier journal author, you have the right to Include the article in a thesis or dissertation (provided that this is not to be published commercially) whether in full or in part, subject to proper acknowledgment; see the Copyright page for more information. No written permission from Elsevier is necessary.

This right extends to the posting of your thesis to your university's repository provided that if you include the published journal article, it is embedded in your thesis and not separately downloadable."

Accessed 2020 https://www.elsevier.com/about/policies/copyright/permissions

This declaration concerns the article entitled:									
High resolution crystal structure of substrate-free human neprilysin									
Publication status (tick one)									
draft manuscript		**Submitted**		**In review**		**Accepted**		**Published**	✓
Publication details (reference)	**Moss S,** Subramanian V & Acharya KR (2018) High resolution crystal structure of substrate-free human neprilysin. J. Struct. Biol. 204, 19–25.								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate contributed to/ considerably contributed to/predominantly executed the… Research design: KRA Methodology design: SM Protein purification and expression of NEP: SM Assessment of activity: SM Crystallisation: SM Structural determination: SM Analysis of data: SM, KRA Wrote the manuscript: SM Prepared the figures: SM Edited the manuscript: SM, VS, KRA								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed	**Stephen Moss**					**Date**	14/04/20		

Journal of Structural Biology 204 (2018) 19–25

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

High resolution crystal structure of substrate-free human neprilysin

Stephen Moss, Vasanta Subramanian, K. Ravi Acharya*

Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK

ARTICLE INFO

Keywords:
Neprilysin
Zinc metalloprotease
Protein structure
Crystallography
Substrate-free

ABSTRACT

Neprilysin is a transmembrane M13 zinc metalloprotease responsible for the degradation of several biologically active peptides including insulin, enkephalin, substance P, bradykinin, endothelin-1, neurotensin and amyloid-β. The protein has received attention for its role in modulating blood pressure responses with its inhibition producing an antihypertensive response. To date, several inhibitor bound crystal structures of the human neprilysin extracellular domain have been determined, but, a structure free of bound inhibitor or substrate has yet to be reported. Here, we report the first crystal structure free of substrate or inhibitor for the extracellular catalytic domain of human neprilysin at 1.9 Å resolution. This structure will provide a reference point for comparisons to future inhibitor or substrate bound structures. The neprilysin structure also reveals that a closed protein conformation can be adopted in protein crystals absent of bound substrate or inhibitor.

1. Introduction

Neprilysin (NEP, EC 3.4.24.11), also known as neutral endopeptidase, neuropeptidase, CD10 and enkephalin, is a transmembrane protease that belongs to the M13 family of zinc metalloproteases. NEP can digest a broad range of peptide substrates including insulin, enkephalin, substance P, bradykinin, endothelin-1, neurotensin and amyloid-β (Kerr and Kenny, 1974; Skidgel et al., 1984; Howell et al., 1995). In addition to the broad substrate specificity, NEP can cleave each substrate at a number of locations but exhibits a preference for the amino side of hydrophobic residues (Hersh and Morihara, 1986).

NEP is located at the plasma membrane and consist of three domains. The short intracellular and transmembrane domains are 27 and 23 residues in length and the third larger extracellular catalytic domain which is 699 residues in length. The extracellular domain contains a large central cavity where the conserved zinc binding motif HEXXH is located. The two histidine residues are responsible for the coordination of the zinc ion, while the glutamate is directly involved in catalysis. This conserved region has been targeted for inhibition to produce an antihypertensive response. To date eight inhibitor bound structures of the human NEP extracellular domain are available in the Protein Data Bank (Oefner et al., 2000; Oefner et al., 2004; Sahli et al., 2005; Oefner et al., 2007; Glossop et al., 2011; Schiering et al., 2016) but a structure free of bound inhibitor or substrate has yet to be reported. Here, we report the first crystal structure of the substrate-free extracellular domain of human NEP.

2. Materials and methods

2.1. Expression and purification

All reagents for expression and purification were sourced from Sigma-Aldrich or Fisher Scientific unless otherwise specified.

A clone of *Pichia pastoris* GS115 expressing the extracellular catalytic domain of human NEP was purchased from Invitrogen. The clone contained human NEP (Tyr51-Trp749) integrated into the *P. pastoris* genome. The cytoplasmic (Gly1-Glu27) and transmembrane (Ile28-Ala50) domains of NEP were absent in the construct to increase the probability of protein crystallisation. The protein of interest was expressed with an N-terminal 6x histidine tag and a secretion peptide sequence.

Expression of the *P. pastoris* clone was completed in accordance with the Invitrogen manual for recombinant protein expression. A glycerol stock of the *P. pastoris* clone was used to inoculate a starter culture containing 25 mL of BMGY (1% Yeast extract, 2% Peptone, 22.968 g L^{-1} K_2HPO_4, 118.048 g L^{-1} KH_2PO_4 pH 6.0, 1.34% Yeast nitrogen base, 4×10^{-5}% Biotin and 1% Glycerol) in a 250 mL baffled flask. The starter culture was incubated overnight at 30 °C. 10 mL of the overnight culture was used to inoculate an expression culture in 2 L baffled flasks containing 250 mL BMGY. The expression culture was incubated at 30 °C for 24 h.

After 24 h the expression culture was harvested at 6000 RCF for 20 min at 19 °C. The supernatant was discarded, and the cell pellet was re-suspended in 250 mL of BMMY (1% Yeast extract, 2% Peptone,

* Corresponding author.
E-mail address: bsskra@bath.ac.uk (K.R. Acharya).

https://doi.org/10.1016/j.jsb.2018.06.004
Received 11 May 2018; Received in revised form 9 June 2018; Accepted 11 June 2018
Available online 12 June 2018
1047-8477/

22.968 g L^{-1} K_2HPO_4, 118.048 g L^{-1} KH_2PO_4 pH 6.0, 1.34% Yeast nitrogen base, 4×10^{-5}% Biotin and 0.5% Methanol) for induction. Re-suspended cells were transferred to a new 2 L baffled flask. The culture was incubated for 72 h at 30 °C. 5 mL of 50% methanol was added to the culture 24 and 48 h.

After 72 h of induction the culture was harvested at 6000 RCF for 20 min at 19 °C. The supernatant was collected and the cell pellet was discarded. Tris and NaCl were added to the supernatant to give final concentrations of 25 mM Tris and 150 mM NaCl. Following the addition of Tris and NaCl the supernatant showed some precipitation which was removed through a 0.22 µm filter (Millipore).

The *P. pastoris* culture supernatant was loaded onto a 5 mL HisTrap affinity column (GE Healthcare Life Sciences) at 2 mL min^{-1} pre-equilibrated with Buffer A (25 mM Tris, 150 mM NaCl, 2 mM $MgCl_2$ pH 7.5). Following loading, the column was washed with 2% Buffer B (25 mM Tris, 150 mM NaCl, 2 mM $MgCl_2$ and 500 mM Imidazole pH 7.5) until the UV trace returned to baseline. Bound NEP was eluted using a gradient of Buffer B between 2 and 100%. 2 mL fractions were collected and analysed using SDS-PAGE with a 10% acrylamide gel to confirm the presence of protein.

Fractions containing protein were concentrated to a volume under 5 mL in a centrifuge at 4000 RCF. For concentration a centrifuge filter unit with a 30 kDa molecule weight cut off (Millipore) was used. The concentrated protein containing sample was then loaded at 5 mL min^{-1} onto a HiPrep 26/10 Desalting column (GE Healthcare Life Sciences) pre-equilibrated with Buffer A. Eluted protein containing fractions were kept at 4 °C for short term storage or −20 °C for longer term storage.

2.2. Enzymatic assay

The enzymatic activity of NEP was assayed by measuring cleavage of the fluorogenic peptide Mca-RPPGFSAFK-(Dnp) (Enzo Life Sciences). A stock solution of 1 mM Mca-RPPGFSAFK(Dnp) dissolved in 100% DMSO was stored at −20 °C. Dilutions of this stock where made using purification Buffer A. Enzymatic assays were conducted using 50 µL of 20 µM Mca-RPPGFSAFK-(Dnp) and 50 µL of 8 nM purified NEP. 50 µL of 8 nM bovine serum albumin (BSA) was used as a negative control. Solutions were manually added to a black 96 well microplate (Greiner Bio-One).

Fluorescence activity assays were performed using a CLARIOstar high-performance microplate reader (BMG LABTECK). Excitation was achieved at 328 nm and emission was detected at 393 nm. Readings were taken in triplicate at 15 s intervals for a length of 45 min.

2.3. Crystallisation

Protein crystallisation was performed with NEP (8 mg/mL) mixed 1:1 with crystallisation buffer 0.2 M KNO_3 and 20% (w/v) PEG3350 (Molecular Dimensions, England). The mixed solution was manually dispensed into 24-well plates (Molecular Dimensions, England) for crystallisation experiments using hanging drop vapour diffusion. Plates were stored at 18 °C and crystals formed within a month. Crystals used for data collection were mounted in a loop and flash-cooled for storage in liquid nitrogen without cryoprotectant.

2.4. Data collection, structure determination and refinement

X-ray diffraction data were collected at the i03 beamline in Diamond Light Source (Didcot, Oxford) at a wavelength of 0.976 Å (12.7 keV) with an oscillation range of 0.1° using a Pilatus3 6 M detector.

Raw X-ray diffraction data were indexed and integrated using DIALS (Waterman et al., 2016). Data were then scaled and merged using AIMLESS within the CCP4 suite (Evans and Murshudov, 2013; Winn et al., 2011). A resolution cut off of 1.9 Å was applied based on statistics generated from AIMLESS (Evans and Murshudov, 2013).

PHASER (McCoy et al., 2007) was used to perform molecular replacement using structure 5JMY (Schiering et al., 2016) on the merged and scaled data from AIMLESS. A single structural solution was provided with a TFZ score of 14.18. After PHASER the structure was refined using REFMAC5 (Murshudov et al., 2011). PHENIX was used towards the end of refinement for its superior handling of protein glycosylation (Adams et al., 2010). Visualisation and model building were performed using COOT (Emsley et al., 2010).

Programs MolProbity (Chen et al., 2010) and PDB validation (Berman et al., 2003) were used to assess any steric and geometry outliers within the modelled structure. The coordination of the active site zinc ion was assessed using the CheckMyMetal server (Zheng et al., 2017) and a composite omit map was generated using Phenix (Terwilliger et al., 2008) to provide a bias free map of the active site. Images were created using either PyMOL (Schrödinger, LLC, 2015) or CCP4mg (McNicholas et al., 2011) and APBS was used for determination of protein cavity electrostatic potential (Baker et al., 2001). RMSD values for main chain superpositions were calculated using LSQ in COOT (Emsley et al., 2010) and cavity volumes were calculated using CASTp (Binkowski et al., 2003).

The final refined structure had R_{work} and R_{free} of 18.99% and 23.54% respectively and was deposited with the Protein Data Bank under code 6GID (Table 1).

Table 1
Neprilysin X-ray data collection and refinement statistics.

Beamline	I03, DLS
Wavelength (Å)	0.976
Crystallographic statistics	
Space group	$P3_221$
Unit cell dimensions	
a, b, c (Å)	107.42, 107.42, 112.32
α, β, γ (o)	90.00, 90.00, 120.00
Resolution range (Å)	93.03–1.90 (1.94–1.90)
R_{merge}	0.128 (1.384)
R_{meas}	0.130 (1.410)
$CC_{1/2}$	1.000 (0.665)
⟨I/σ (I)⟩	25.3 (3.7)
Completeness (%)	100 (100)
No. observed reflections	3,193,442 (201,035)
No. unique reflections	59,422 (3,790)
Multiplicity	53.7 (53.0)
Refinement statistics	
R_{work}/R_{free} (%)	18.99/23.54
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (°)	0.894
Ramachandran statistics	
Favoured (%)	98
Allowed (%)	2
Outliers (%)	0
Wilson B-factor ($Å^2$)	30.8
Average B-factors ($Å^2$)	
Protein	34.0
Water	38.1
Zinc	24.9
Phosphate	37.1
Carbohydrate	54.5
No. of atoms	
Protein	5599
Water	463
Carbohydrate	56
PDB code	**6GID**

Values in parentheses are for highest-resolution shell.

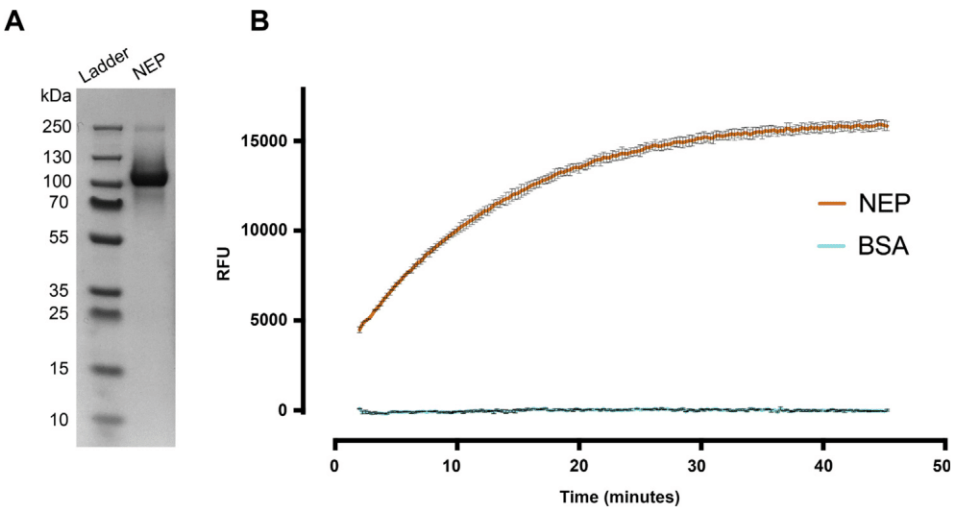


Fig. 1. (A) SDS-PAGE of purified NEP sample. (B) Enzymatic assay for NEP measuring relative fluorescent units (RFU) over time. Fluorescent substrate Mca-RPPGFSAFK (Dnp) (10 μM), NEP (8 nM) and BSA (8 nM) were used. Each assay was conducted in triplicate and the average has been plotted with standard error bars.

3. Results and discussion

Here we report the first substrate-free crystal structure for the extracellular catalytic domain of human NEP (Gly54-Trp749). The structure was determined in the trigonal space group $P3_221$ with a single molecule in the asymmetric unit to a resolution of 1.9 Å. The protein used for this structural study was expressed and purified in *P. pastoris* and has been shown to be catalytically active (Fig. 1).

The reported structure was refined to give final R_{work} and R_{free} values of 18.99% and 23.54% respectively. Overall, 98% of residues are within the preferred region of the Ramachandran plot with 2% allowed and no outliers. The details of data collection and refinement statistics are presented in Table 1.

3.1. General features

The NEP extracellular domain is ellipsoid in shape with a long and short axis of 86 Å and 60 Å respectively. The secondary structure of NEP is mainly alpha helical in nature and can be further separated into two large lobe-like subdomains connected by a smaller linker region. The linker region is formed of four alpha helical segments distributed non-consecutively in the protein sequence (Fig. 2). Subdomain 1 consists of N-terminal residues while subdomain 2 consists of C-terminal residues. Together, subdomains 1 and 2 enclose a large central cavity.

Four sites of *N*-linked glycosylation have been confirmed at positions N144, N284 and N324 in subdomain 2 and position N627 in subdomain 1. Electron density allowed a single *N*-acetylglucosamine residue to be modelled at each of these sites.

The catalytic site of NEP, located within the cavity surface of subdomain 1, is based around a zinc ion and conserved active site motif HEXXH. The zinc ion is coordinated by conserved residues H583 (2.02 Å) and H587 (2.10 Å) and additional coordination is provided by E646 (1.93 Å). The third conserved residue, E584, is involved in the catalytic mechanism and adopts a position above the zinc ion completing coordination sphere.

The coordinated zinc ion and surrounding residues creates a small binding pocket with subsites S1, S1′ and S2′ (Fig. 3). Inhibitors designed for NEP have revealed characteristics about the subsites. The S1 site has low impact on the overall affinity for binding and displays relaxed specificity (Oefner et al., 2000). S1′ subsite forms a hydrophobic pocket that exhibits specificity for large hydrophobic and aromatic side chains (Tiraboschi et al., 1999). This accounts for the preference for NEP to cleavage sites at the amino side of hydrophobic residues. The S2′ subsite displays relaxed specificity and can accommodate bulky side chains (Dion et al., 1997).

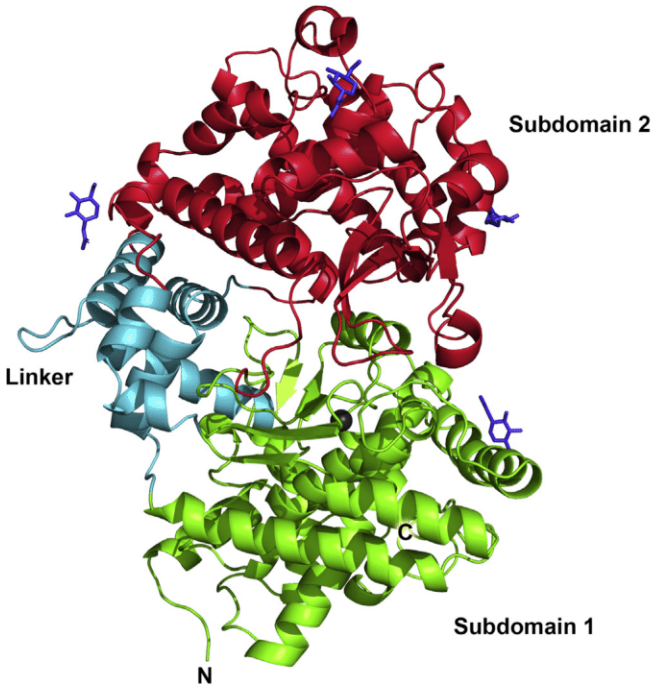


Fig. 2. Three-dimensional ribbon diagram of the NEP extracellular domain. The extracellular domain is further split into two subdomains coloured green for subdomain 1, red for subdomain 2 and cyan for the linker region. Observed glycan residues are displayed in purple and the zinc ion is displayed in grey.

3.2. Modelling active site density

In addition to residues H583, H587 and E646, a ligand is required to create an acceptable coordination sphere for the catalytic zinc ion. In the structure reported here, this is provided by a phosphate ion (2.07 Å and 2.36 Å) carried through from the *P. pastoris* expression growth media. The final refined structure shows the modelled phosphate ion has a good fit to the electron density (Fig. 4A). A composite omit map was used to generate bias-free electron density for the region surrounding the catalytic zinc ion (Fig. 4B). The omit map supported the placement of the phosphate ion, with good density present above the zinc ion. The full coordination sphere was evaluated using the CheckMyMetal web server (Zheng et al., 2017) to confirm the acceptable coordination interactions were modelled.

The presence of a ligand coordinating the catalytic zinc ion has been reported in other native zinc metalloprotease crystal structures that

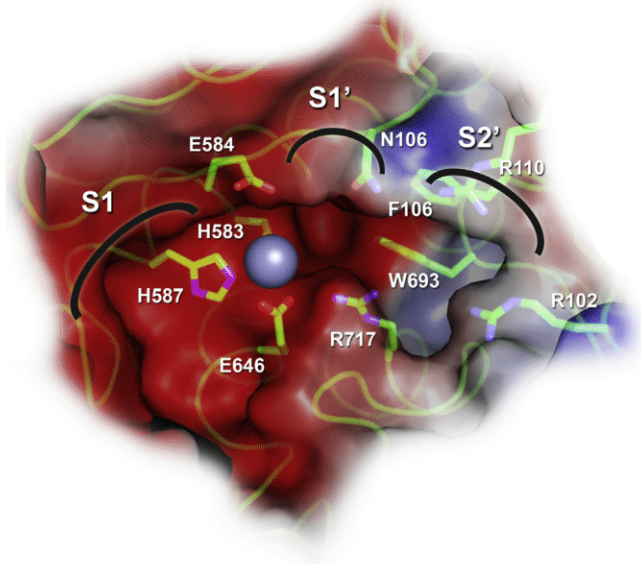


Fig. 3. NEP active site binding pocket labelled with subsites and residues.

share the HEXXH motif. In angiotensin-converting enzyme (ACE, 1O8A) an acetate is responsible for coordinating the zinc atom (Natesh et al., 2003). Other structures including insulin-degrading enzyme (IDE, 2JG4) (Im et al., 2007) and thermolysin (1TLX and 2TLX) (English et al., 1999), have water molecules coordinating the zinc ion. However, in these structures the water molecule does not completely account for the electron density above the zinc ion and the coordination distances between water molecule and the zinc ion are greater than acceptable lengths (Zheng et al., 2017).

It is clear that the molecule responsible for the coordination of the catalytic zinc ion has an area of density that is difficult to model with certainty in native zinc metalloprotease structures. It is likely that a range of ligands or water molecules are responsible for coordinating the zinc ion across the different protein molecules present within the protein crystal. As a result, the density describes an average of these ligands and water molecules. In the structure reported here, a phosphate ion has been modelled to coordinate the zinc ion based on the evidence present above.

The structure reported here contains additional weaker density in the S1′ and S2′ subsites. That weaker density has been modelled as a polyethylene glycol fragment (PEG) acquired from the crystallisation buffer. Occupancy of the PEG fragment was refined using Phenix to 0.95. The fragment has several favourable contacts to the phosphate ion but has limited interactions to active site residues with no contacts closer than 3.15 Å in length.

The native ACE, IDE and thermolysin structures also exhibit additional weak density in the active site. It is likely that a range of small molecules present in the crystallisation buffers show affinity for the binding site accounting for weak density in the region. Due to the presence of a PEG molecule and phosphate ion we describe this structure as substrate-free opposed to native.

3.3. Active site flexibility

To date, eight inhibitor bound crystal structures of NEP extracellular domain are currently available in the Protein Data Bank. These structures were superimposed onto the substrate-free NEP structure and active site residues were visually inspected for conformational changes. Two inhibitor bound structures were selected to accurately describe all conformational changes that were present across the eight available structures. These two structures are compared to substrate-free NEP below.

First, the substrate-free crystal structure is compared to crystal structure 5JMY (Schiering et al., 2016) which contains inhibitor LBQ657. Conformational changes in residues F106 and W693 can be seen between our substrate-free structure and 5JMY (Fig. 5A). W693 displays a large shift that increases the volume of the S1′ pocket and allows the biphenyl group of LBQ657 to occupy the site. The shift in W693 causes rotation in F106 altering the S2′ subsite. The second comparison is to crystal structure 1DMT (Oefner et al., 2000) which contains inhibitor phosphoramidon. Conformational changes in R110 form the interface of the S2′ subsite and accommodate for the bulky side chain of phosphoramidon (Fig. 5B).

The flexibility in residues listed above helps explain how NEP can accommodate a number of different peptide substrates at different locations with variability in the specific amino acid that interacts with the cleavage site. W693, F106 and R110 are the only active site amino acids that show significant motions when compared to the eight published inhibitor bound crystal structures.

3.4. Conformational flexibility

The crystal structure reported here for the NEP extracellular domain is in a closed conformation. Two small solvent accessible entrances to the large cavity are located at the front and rear of the protein measuring approximately 8 Å and 6 Å at their narrowest points (Fig. 7A). These entrances are too small for the entry of substrate or inhibitor indicating NEP must undergo conformational changes to adopt an open conformation.

Published inhibitor bound structures all report NEP in closed conformations. Superpositions of these structures with substrate-free NEP (Fig. 6), and RMSD values calculated for the superpositions (Table 2), indicate the close conformational similarity. This reveals binding of substrate or inhibitor is not required to stabilise the closed conformation in the NEP crystal structure. It is unclear if the protein only opens when interacting with substrate or whether the closed conformation is a result of the protein existing in a crystal structure.

In a closed conformation NEP's large central cavity has an approximate volume of 5000 $Å^3$ which is a smaller volume than many of the peptides it is responsible for cleaving (Malito et al., 2008). Previous

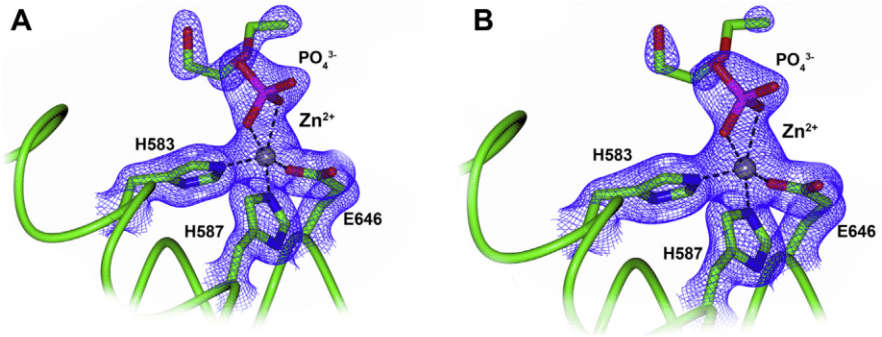


Fig. 4. The active site zinc ion is coordinated by residues H583, H587, E646 and a phosphate ion. An additional PEG fragment is located in proximity to the phosphate ion. Electron density maps of the coordination sphere of the refined structure is shown. (A) 2mFo-DFc and (B) composite omit map. Both maps are shown to a contour level of 1.5σ.

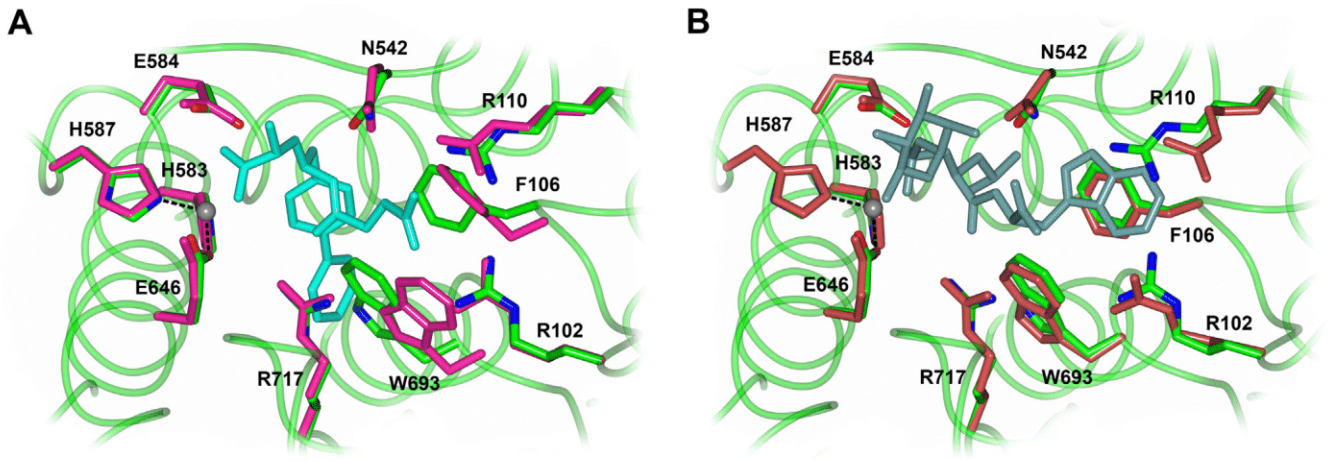


Fig. 5. Substrate-free NEP active site (green) is superimposed with 5JMY (Schiering et al., 2016) (pink) (B) and 1DMT (Oefner et al., 2000) (brick) (C). Inhibitors LBQ657 and phosphoramidon are shown in cyan and jade, with the zinc ion shown in grey.

Fig. 6. Mainchain superpositions of substrate-free NEP with published structures: 5JMY (magenta), 2YB9 (light grey), 2QPJ (violet), 1Y8J (lime green), 1R1H (purple), 1R1I (grey), 1R1J (blue) and 1DMT (cyan).

studies have revealed amyloid-β, a substrate of NEP, is cleaved at a number of locations within the centre of the polypeptide chain (Sexton et al., 2012; Webster et al., 2014). Cleavage in this manner is a characteristic feature seen in other zinc metalloproteases including insulin degrading enzyme (IDE) which has a similar the two lobe-like domain structure with large central cavities.

In order for substrate cleavage to occur, it is likely that large movements in the subdomains are required for substrate access into the active site. Structures of IDE have revealed that a separation of the two lobe-like domains allows entry of substrate before a closed

Table 2
RMSD values from published inhibitor bound NEP structures superimposed onto the substrate-free structure. RMSD values were calculated in COOT using a least squares comparison for mainchain residues 54–749.

PDB code	RMSD	Reference
5JMY	0.426	Schiering et al. (2016)
2YB9	0.354	Glossop et al. (2011)
2QPJ	0.452	Oefner et al. (2007)
1Y8J	0.421	Sahli et al. (2005)
1R1H	0.385	Oefner et al. (2004)
1R1I	0.484	Oefner et al. (2004)
1R1J	0.421	Oefner et al. (2004)
1DMT	0.340	Oefner et al. (2000)

conformation is adopted completely engulfing the substrate into the central cavity (McCord et al., 2013). Given the smaller size of the NEP cavity it is unlikely that the protein completely engulfs its substrate, but it is clear that motion is required for access to the active site. The exact mechanism will remain elusive until peptide bound structures of NEP are determined.

The electrostatic surface charge of the cavity has been analysed using APBS (Fig. 7B). A striking electrostatic difference between the top and bottom surfaces of the cavity is present. The top surface, contributed by subdomain 2, displays a large area of positive electrostatic potential. The opposite is observed for the bottom surface, contributed by subdomain 1, where a large negative electrostatic potential is present in and in proximity to the catalytic site. It is also observed that the front entrance to the cavity possesses a strong negative electrostatic potential, while the rear entrance possesses a strong positive electrostatic potential. These electrostatic features may be critical for guidance of substrate to the active site and important for substrate binding.

4. Conclusion

In this report we have presented a substrate-free crystal structure for the extracellular domain of human NEP. The active site of this structure has been compared to other available inhibitor bound structures revealing flexibility in active site residues W693, F106 and R110. Flexibility in these residues is critical for allowing the active site to accommodate a variety of substrates binding at a range of positions. Additionally, we also note the central cavity of NEP has striking electrostatic potential differences between the top and bottom side. The function of these differences is unknown but are likely involved in substrate interaction.

The crystal structure of substrate-free NEP reveals that a closed

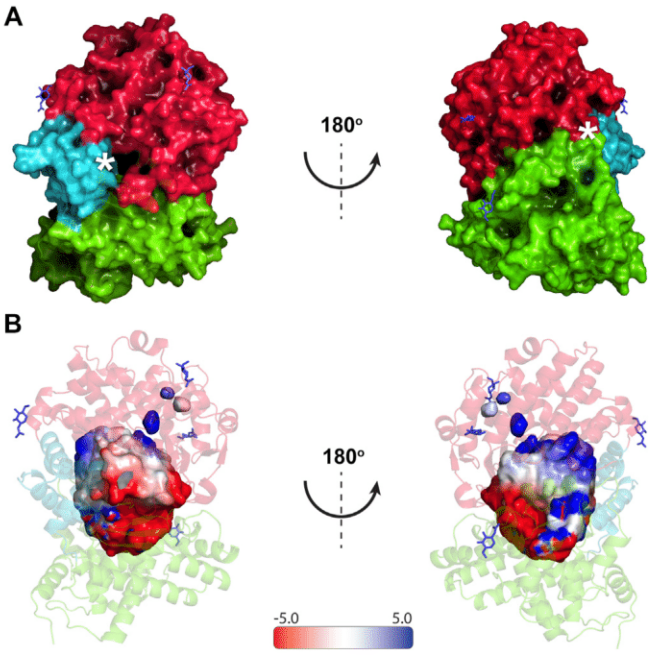


Fig. 7. (A) Molecular surface representation of front (left) and rear (right) of NEP and coloured green for subdomain 1, red for subdomain 2 and cyan for linker region. Small solvent entrances are indicated with an asterisk. (B) Molecular surface representation of front (left) and rear (right) for the central cavity of NEP. Electrostatic charge has been calculated with APBS. The electrostatic potential is measured in eV with scale bar indicated within the figure.

protein conformation can be retained in the absence of bound substrate or inhibitor. The mechanism for substrate entrance into the active site remains unclear, but clues provided by IDE suggest large domain shifts may be required for substrate entry. The structure reported here provides a key reference point for any future structural studies investigating substrate binding and conformational movement.

Acknowledgements

We thank Diamond Light Source for access to beamline I03 (proposal mx17212) that resulted in the data presented here. S.M is supported by a post-graduate studentship from the Alzheimer's Society (UK) awarded to K.R.A. and V.S. [grant number – 286 (AS-PhD2015b-006)]. This research made use of the Balena High Performance Computing (HPC) Service at the University of Bath – UK.

Author contributions

S.M performed all the experiments, analysed the data and wrote the manuscript. V.S. supervised the study and edited the manuscript. K.R.A. conceptualised and supervised the study, analysed the data and edited the manuscript.

Additional information

Accession code: The atomic coordinates and structure factors (code 6GID) have been deposited in the Protein Data Bank (http://www.pdb.org).

Competing financial interests

The authors declare no competing financial interests.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. Phenix: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 213–221.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. 98, 10037–10041.

Berman, H., Henrick, K., Nakamura, H., 2003. Announcing the worldwide Protein Data Bank. Nat. Struct. Mol. Biol. 10, 980.

Binkowski, T., Naghibzadeh, S., Liang, J., 2003. CASTp: computed atlas of surface topography of proteins. Nucleic Acids Res. 31, 3352–3355.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 12–21.

Dion, N., Cohen, P., Crine, P., Boileau, G., 1997. Characterisation of neprilysin (EC 3.4.24.11) S2′ subsite. FEBS Lett. 411, 140–144.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 486–501.

English, A.C., Done, S.H., Caves, S.L., Groom, R.C., Hubbard, E.R., 1999. Locating interaction sites on proteins: the crystal structure of thermolysin soaked in 2% to 100% isopropanol. Proteins 37, 628–640.

Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? Acta Crystallogr. Sect. D Biol. Crystallogr. 69, 1204–1214.

Glossop, M.S., Bazin, R.J., Dack, K.N., Fox, D.N.A., MacDonald, G.A., Mills, M., Owen, D.R., Phillips, C., Reeves, K.A., Ringer, T.J., Strang, R.S., Watson, C.A.L., 2011. Synthesis and evaluation of heteroarylalanine diacids as potent and selective neutral endopeptidase inhibitors. Bioorg. Med. Chem. Lett. 21, 3404–3406.

Hersh, L.B., Morihara, K., 1986. Comparison of the subsite specificity of the mammalian neutral endopeptidase 24.11 (Enkephalinase) to the bacterial neutral endopeptidase Thermolysin. J. Biol. Chem. 261, 6433–6437.

Howell, S., Nalbantoglu, J., Crine, P., 1995. Neutral endopeptidase can hydrolyze β-amyloid (1–40) but shows no effect on β-amyloid precursor protein metabolism. Peptides 16, 647–652.

Im, H., Manolopoulou, M., Malito, E., Shen, Y., Zhao, J., Neant-Fery, M., Sun, C.Y., Meredith, S.C., Sisodia, S.S., Leissring, M.A., 2007. Structure of substrate-free human insulin-degrading enzyme (IDE) and biophysical analysis of ATP-induced conformational switch of IDE. J. Biol. Chem. 282, 25453–25463.

Kerr, M.A., Kenny, A.J., 1974. The purification and specificity of a neutral endopeptidase from rabbit kidney brush border. Biochem. J. 137, 477–488.

Malito, E., Hulse, R.E., Tang, W.J., 2008. Amyloid-β degrading cryptidases: insulin degrading enzyme, presequence peptidase, and neprilysin. Cell. Mol. Life Sci. 65, 2574–2585.

McCord, L.A., Liang, W.G., Dowdell, E., Kalas, V., Hoey, R.J., Koide, A., Koide, S., Tang, W.-J., 2013. Conformational states and recognition of amyloidogenic peptides of human insulin-degrading enzyme. Proc. Natl. Acad. Sci. 110, 13827–13832.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674.

McNicholas, S., Potterton, E., Wilson, K.S., Noble, M.E.M., Emsley, P., Murshudov, G.N., Cohen, S., Perrakis, A., Noble, M., 2011. Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 386–394.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 355–367.

Natesh, R., Schwager, S.L., Sturrock, E.D., Acharya, K.R., 2003. Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. Nature 421, 551–554.

Oefner, C., D'Arcy, A., Hennig, M., Winkler, F.K., Dale, G.E., 2000. Structure of human neutral endopeptidase (neprilysin) complexed with phosphoramidon. J. Mol. Biol. 296, 341–349.

Oefner, C., Pierau, S., Schulz, H., Dale, G.E., 2007. Structural studies of a bifunctional inhibitor of neprilysin and DPP-IV. Acta Crystallogr. Sect. D 63, 975–981.

Oefner, C., Roques, B.P., Fournie-Zaluski, M.C., Dale, G.E., 2004. Structural analysis of neprilysin with various specific and potent inhibitors. Acta Crystallogr. Sect. D 60, 392–396.

Sahli, S., Frank, B., Schweizer, W.B., Diederich, F., Blum-Kaelin, D., Aebi, J.D., Böhm, H., Oefner, C., Dale, G.E., 2005. Second-generation inhibitors for the metalloprotease neprilysin based on bicyclic heteroaromatic scaffolds: synthesis, biological activity, and X-ray crystal-structure analysis. Helv. Chim. Acta 88, 731–750.

Schiering, N., D'arcy, A., Villard, F., Ramage, P., Logel, C., Cumin, F., Ksander, G.M., Wiesmann, C., Karki, R.G., Mogi, M., 2016. Structure of neprilysin in complex with the active metabolite of sacubitril Sci. Rep. 6, 27909. http://dx.doi.org/10.1038/srep27909.

Schrödinger, L., 2015. The PyMOL molecular graphics system. Version 1, 8.

Sexton, T., Hitchcook, L.J., Rodgers, D.W., Bradley, L.H., Hersh, L.B., 2012. Active site mutations change the cleavage specificity of neprilysin. PLoS One 7, 1–10.

Skidgel, R.A., Engelbrecht, S., Johnson, A.R., Erdös, E.G., 1984. Hydrolysis of substance P and neurotensin by converting enzyme and neutral endopeptidase. Peptides 5, 769–776.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Adams, P.D., Read, R.J., Zwart, P.H., Hung, L.W., 2008. Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. Acta Crystallogr. Sect. D 64, 515–524.

Tiraboschi, G., Jullian, N., Thery, V., Antonczak, S., Fournie-Zaluski, M.-C., Roques, B.P., 1999. A three-dimensional construction of the active site (region 507–749) of human neutral endopeptidase (EC.3.4.24.11). Protein Eng. Des. Sel. 12, 141–149.
Waterman, D.G., Winter, G., Gildea, R.J., Parkhurst, J.M., Brewster, A.S., Sauter, N.K., Evans, G., Sauter, K., Winter, G., Toyokawa, H., Wagner, A., 2016. Diffraction-geometry refinement in the DIALS framework. Acta Crystallogr. Sect. D Biol. Crystallogr. 39, 558–575.
Webster, C.I., Burrell, M., Olsson, L.-L., Fowler, S.B., Digby, S., Sandercock, A., Snijder, A., Tebbe, J., Haupts, U., Grudzinska, J., Jermutus, L., Andersson, C., 2014. Engineering neprilysin activity and specificity to create a novel therapeutic for Alzheimer's Disease. PLoS One 9, e104001.
Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., IUCr, 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 235–242.
Zheng, H., Cooper, D.R., Porebski, P.J., Shabalin, I.G., Handing, K.B., Minor, W., 2017. CheckMyMetal: a macromolecular metal-binding validation tool. Acta Crystallogr. Sect. D 73, 223–233.

3.5 Post-paper conclusion

In this chapter a method for expression, purification and crystallisation of NEP has been established. Using the methods described, determination of the substrate-free NEP structure has been possible. That structure has provided important information regarding the active site in an unbound form. It is now clear that the active site is static in nature with inhibitor binding inducing only minimal movements involving three residues W693, F106 and R110. With a substrate-free structure determined and a methodology in place for production of NEP crystals, it is now possible to shift focus to structurally investigate the substrate binding interactions of NEP. This topic will be described in Chapter 5.

Another avenue for further investigation is in understanding the mechanism for substrate entrance into the central cavity of NEP. The possibility that inhibitor binding induced the closed conformation in NEP crystal structures can now be ruled out. With that in mind two possibilities remain. First, the closed form is the favourable conformation NEP adopts in the absence of bound substrate. Second, the closed conformation could be driven by crystal packing forces in the crystallisation of NEP. Chapter 4 investigates this further through the use of small angle scattering as a method to determine the conformations of NEP present in solution.

Chapter 4 – Conformations of neprilysin in solution

4.1 Foreword

In Chapter 3 the substrate-free crystal structure of NEP was determined to be in a closed conformation. To validate this structure, and investigate the conformational dynamics of NEP in solution, the technique of small angle X-ray scattering (SAXS) was employed. This short chapter is formatted as a conventional thesis chapter and covers the analysis of NEP using SAXS.

4.2 Introduction

4.2.1 Small angle X-ray scattering

Small angle X-ray scattering (SAXS) is an experimental technique which provides structural information in the range of 1-100 nm. Compared to X-ray crystallography, SAXS is a low-resolution technique, providing information on protein size and shape. As such, SAXS is regularly used to determine protein oligomeric state and, in the case of multi-protein assembles, spatial organisation. SAXS can also be used to validate protein structures determined using X-ray crystallography and exclude the possibility that observed conformations are a result of the local crystal packing environment. While these applications for SAXS are useful, its real strength lies in its power to investigate the dynamics and conformation changes (in solution state) that occur in proteins over time. Frequently SAXS is used in combination with molecular dynamics (MD) simulations to gain comprehensive knowledge of structural motions that are supported through experimental data (Boldon *et al.*, 2015).

4.2.2 SAXS sample preparation and data collection

Preparation of protein samples for a SAXS experiment is relatively simple with the only essential requirement being that protein particles are monodisperse in solution. Other sample parameters can be varied, such as protein concentration, in order to provide better signal-to-noise ratio. Once a monodispersed protein solution has been prepared it is possible to begin data collection, the aim of which is to record the scattering produced when a monochromatic X-ray beam is passed through a protein solution or a buffer blank. Due to the random orientations of particles in solution the scattering pattern observed is

isotropic and a rotational averaging of scattered X-rays (Figure 4.1). Once scattering data is collected buffer subtraction can be used to identify the scattering components contributed by protein particles.

Two main methods exist to conduct SAXS data collection, these are batch mode and HPLC mode. In batch mode, protein and buffer samples are separately loaded into a quartz capillary before X-ray exposure. In HPLC mode, also called size exclusion chromatography SAXS (SEC-SAXS), the protein sample is passed through a size exclusion column and the eluted solution is continuously exposed to X-rays. Whilst HPLC mode takes substantially longer than batch mode, there are two benefits. First, high quality buffer subtraction can be attained as the contribution of buffer scattering can be assessed from the buffer exposed to X-rays before protein is eluted. Second, HPLC mode guaranties that protein solution will be monodispersed.

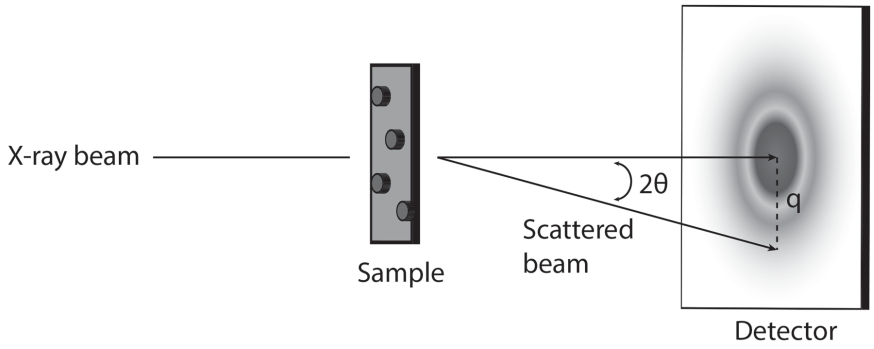


Figure 4.1, SAXS experiment schematic. A monochromatic X-ray beam is passed through a protein sample. The X-ray scattering is recorded on a detector to a maximal angle of 2ɵ. Scattering data can then be reduced, and structural analysis performed.

4.2.3 Data reduction and analysis

After SAXS data has been collected indexing can commence. Here the scattering vector *q* (also referred to as *s*) (Equation 4.1) is calculated from the scattering angle (θ) and the wavelength of the X-ray beam (λ).

$$q = \frac{4\pi \sin(\theta)}{\lambda} \qquad \textit{Equation 4.1}$$

Scattering vectors are then assigned a scattering intensity *I(q)* which is composed of two components: particle form factors *F(q)* and structure factors *S(q)* (Equation 4.2). The form factors are related to the intra-particle contacts and provided information regarding the shape and size of the particles. Structure factors are related to interparticle attraction or repulsion interactions and, if particles are monodispersed, *S(q)* = 1. One of the most common issues in SAXS is that aggregation changes results in *S(q)* ≠ 1. This affects the recorded measured scattering intensity and altering approximations of form factors.

$$I(q) = F(q)\,S(q) \qquad \textit{Equation 4.2}$$

Once values for the scattering vector and intensity have been determined, the spatial variations in electron density between particles and solvent, also referred to as contrast, can be calculated (Kikhney and Svergun, 2015). The contrast can be used to approximate several particle parameters including radius of gyration (R_g), maximum particle dimension (D_{max}), molecular weight and flexibility. These values can in turn be used to determine oligomeric state, protein complex organisation and predict an overall low-resolution protein model or 'molecular envelope'.

The first stage in data analysis is to integrate the recorded scattering vectors and intensities onto a 1-dimensional SAXS profile (log *I(q)* by *q*) (Figure 4.2). The 1-dimensional profile contains structural information on the protein molecule, with low *q* scattering revealing low resolution information while higher *q* scattering provides higher resolution information. In order to extract structural information from the SAXS profile and determine structural parameters including R_g, D_{max}, molecular weight and flexibility the 1-dimensional profile must be transformed into a series of other plots.

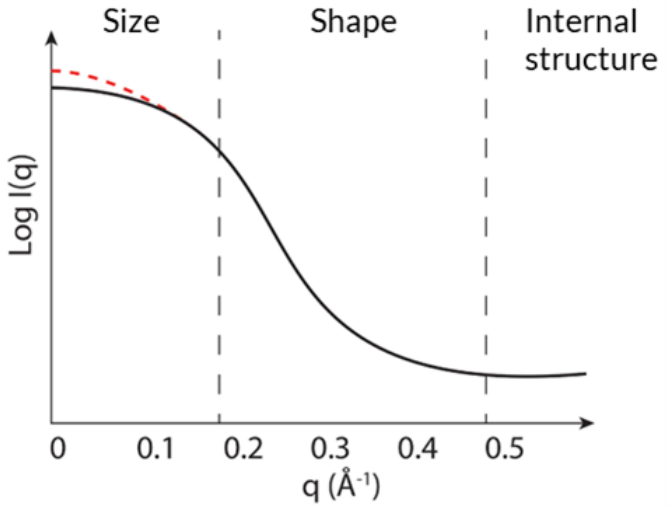


Figure 4.2, 1-dimensional SAXS profile. SAXS data can be integrated and plotted onto a graph of log I(q) by q. Data at low q provides low resolution size information. High q data give higher resolution information about internal structure.

The first of these manipulations is frequently the Guinier plot which uses low q data (q x R_g < 1.3) to estimate R_g (Figure 4.3a). Low q data can be described by a linear equation (Equation 4.3) where the gradient is R_g and the intercept is the intensity at zero scattering angle ($I(0)$). If the low q data cannot be fitted to a linear equation it is often the first indication that there is a problem with the sample. Commonly this occurs when the $S(q) \neq 1$ because there is some form of attraction/aggregation making the data inappropriate for analysis.

$$lnI(q) = lnI_0 - \frac{{R_g}^2}{3}q^2 \qquad \textit{Equation 4.3}$$

The second manipulation in SAXS analysis is to convert the data into real space using an indirect Fourier transform or the 'pair distance distribution function' $P(r)$. Like the Guinier plot a $P(r)$ plot (Figure 4.3b) can also estimate R_g, however R_g is calculated using all the data instead of just low q scattering. This provides a better estimate less affected by $S(q)$. A $P(r)$ plot also provides an estimation of D_{max} at the x-axis intercept.

R_g and D_{max} are the best-known parameters that can be calculated from SAXS but other parameters including mass and particle flexibility can also be estimated. Several methods are used to calculate the molecular weight of a protein from SAXS data. If the initial scattering intensity *I(0)* is known then an estimate of electron number can be made from which molecular weight can be determined. Alternatively, the average protein density of 1.37 g/cm^3 can be used with the calculated volume to estimate molecular weight.

Kratky analysis (Figure 4.3c) allows approximation of the flexibility in system and can be used to determine if a particle is globular, partially folded or unfolded. It is particularly useful to visualise structural changes that alter a particles compactness and or flexibility. Frequently a normalised Kratky plot (Figure 4.3d) provides an even more convenient way to visualise conformation movement/s by removing information on particle size but retaining information on shape. For a globular particle, there will be a peak maximum at √3 regardless of particle size or composition. Divergence from this indicates flexibility or asymmetry (Receveur-Brechot and Durand, 2012).

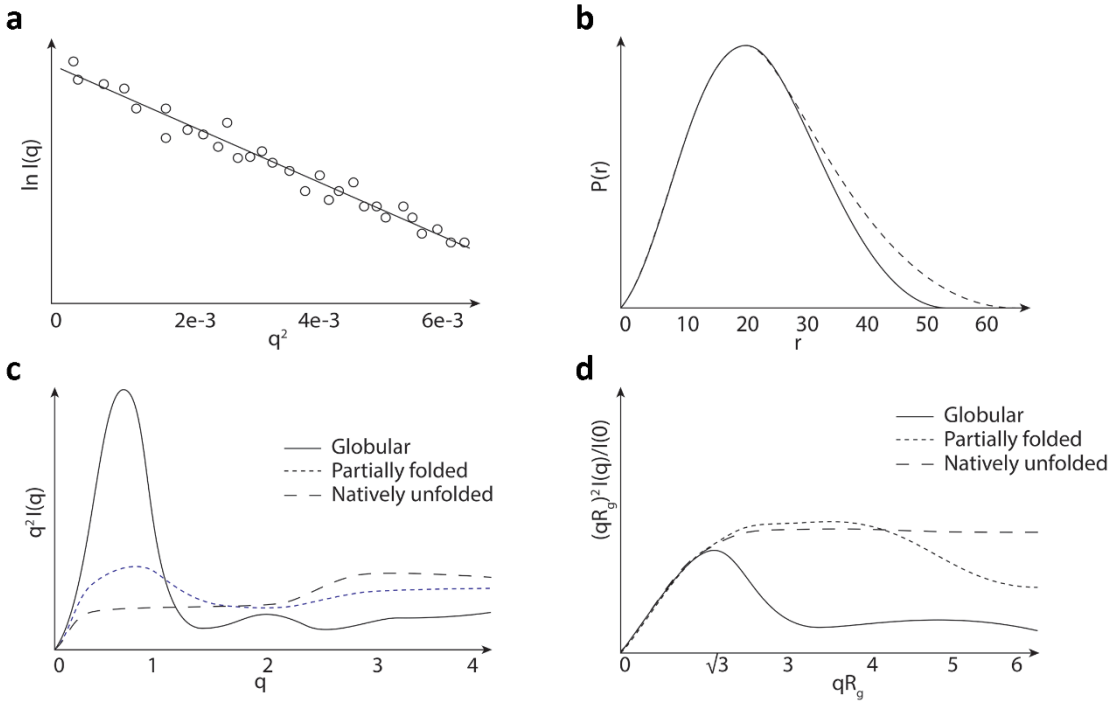


Figure 4.3, Example SAXS data analysis and plots. (a) Guinier plot is used to approximate the R_g of a particle using just low q data that can be described by a linear equation. (b) P(r) plot is an indirect Fourier transform converting SAXS data into real space. P(r) allows estimation of R_g using all data and calculation of D_{max}. (c) Kratky plot provides an approximation of particle flexibility and can be used to observe conformation changes. (d) Normalised Kratky

plots show changes in particle shape independent of particle size and can be a more convenient way to observe structural changes (Kikhney and Svergun, 2015).

Commonly the last of manipulation of SAXS data is the generation of 3-dimensional model in the process of envelope reconstruction. Dammif is a well-known programme for envelope reconstruction and generates a random particle model from spheres. During refinement spheres are added and rejected until the theoretical scattering pattern is similar to the pattern that was experimentally observed. The programme Damaver averages different Dammif runs that have started from random seeds to improve the accuracy of the models built. While envelope reconstruction is a nice visual way to represent SAXS data, envelopes have issues surrounding local minima and should not be over interpreted.

4.2.4 Neprilysin in solution

In Chapter 3 the crystal structure of substrate-free NEP was determined in a closed conformation (Figure 4.4). The technique of SAXS provides a method to confirm that this conformation exists in solution and is not a crystallographic artifact. SAXS may also provide some indication of the conformational flexibility exhibited by NEP in solution. In order to degrade peptides, it is clear that some conformational motion is required to allow substrates into the active site within the central cavity. From the NEP crystal structure, it is possible to identify a likely mode of conformational motion between non-covalently linked regions of subdomains (Figure 4.4). Determination of scattering parameters R_g and D_{max} will allow investigation into any conformational changes that occur in SAXS experiments. Furthermore, it may be possible to shift the equilibrium of NEP to an open or closed conformation through variation in protein buffer conditions.

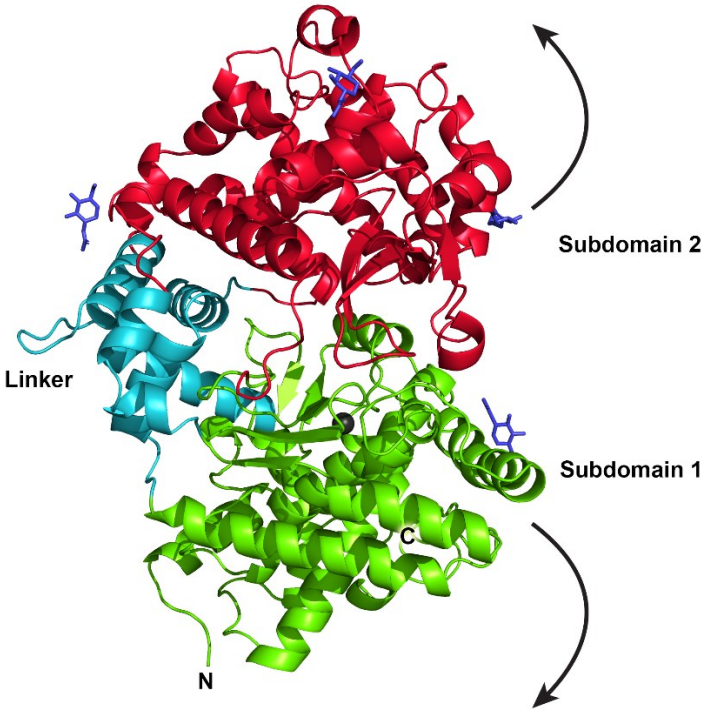


Figure 4.4, Crystal structure of human NEP extracellular domain. The extracellular domain can be separated into two subdomains attached by a linker region. Subdomains act to enclose a large central cavity occluded from substrates in the closed conformation. Arrows indicate the likely mode of domain opening to allow substrate entrance.

4.3 Methods

All reagents for expression and purification were sourced from Sigma-Aldrich or Fisher Scientific unless otherwise specified. General methods were followed as described in Chapter 3 (3.2.1).

4.3.1 Buffers

NEP pH 5.5 SAXS buffer	25 mM MES, 150 mM NaCl, 2 mM $MgCl_2$ (pH 5.5, pH adjusted with NaOH).
NEP pH 7.5 SAXS buffer	25 mM Trizma base, 150 mM NaCl, 2 mM $MgCl_2$ (pH 7.5, pH adjusted with HCl).
NEP pH 9.5 SAXS buffer	25 mM CHES, 150 mM NaCl, 2 mM $MgCl_2$ (pH 9.5, pH adjusted with HCl).
NEP low-salt SAXS buffer	25 mM Trizma base, 150 mM NaCl, 2 mM $MgCl_2$ (pH 7.5, pH adjusted with HCl).
NEP high-salt SAXS buffer	25 mM Trizma base, 500 mM NaCl, 2 mM $MgCl_2$ (pH 7.5, pH adjusted with HCl).

Table 4.1, Buffers used during SAXS analysis of NEP.

4.3.2 Data collection

SAXS experiments were conducted at DLS on beamline B21 using a PILATUS-6M detector. Experiments were set up in batch or HPLC mode.

Batch mode: Protein samples were concentrated to 30 µl and 2.5 mg/ml in SAXS buffer pH 7.5 using a 30 kDa molecular weight cut-off concentrator (Millipore). Buffer exchange was required to substitute NEP into SAXS buffer pH 5.5 and SAXS buffer 9.5. This was achieved by adding 470 µl exchange buffer to the 30 µl sample. This was then concentrated down to 30 µl before repeating the process three further times. The final flow-through was collected and used as a SAXS blank for accurate buffer substation.

HPLC mode: Proteins samples were concentrated to 60 µl and 2.5 mg/ml in either NEP low-salt SAXS buffer or NEP high-salt SAXS buffer. Samples were loaded onto a Superdex200 column (GE Healthcare) pre-equilibrated with their respective SAXS buffer. Column flow rate was 0.075 mL min^{-1} and eluted protein was passed through an X-ray beam at 12400 eV.

4.3.3 Data reduction and analysis

SAXS data processing was done within the ATSAS suite using programme Primus (Franke *et al.,* 2017). Guinier and *P(r)* plots were used for estimation of R_g and D_{max}. Predicted scattering data for the substrate-free NEP crystal structure was conducted using Crysol (Franke *et al.,* 2017) and PDB entry 6GID (Moss *et al.,* 2018). The substrate-free NEP crystal structure was altered using glycam (Woods Group, 2019) to add sugar residues so there were 12 monosaccharide residues attached to each of the four glycosylation sites. This glycosylation was used as an approximation from the mass spectrometry data calculated in Chapter 3. X^2 values of the correlation between predicted and experimentally calculated data was determined in Crysol. Kratky plots to illustrate the correlation between predicted and experimentally calculated data were generated using R (R Core Team, 2019).

4.4 Results

4.4.1 Data collection and processing

NEP was expressed and purified in accordance with the methods described in Chapter 3. This protein was buffer exchanged into varied pH and salt conditions to determine the effect on NEP conformation in a soluble environment. For pH variation, batch mode SAXS was conducted using NEP at 2.5 mg/mL in protein buffer solutions of pH 5.5, 7.5 and 9.5. Scattering data were collected, indexed and then plotted onto 1-dimensional profiles (Figure 4.5a, c and e) before performing a Guinier analysis to estimate R_g values (Figure 4.5b, 4d, and 4f). All datasets displayed linear low q data which indicated that protein samples were appropriately monodisperse. The calculated R_g values in addition to D_{max} values calculated from *P(r)* analysis are shown in Table 4.2.

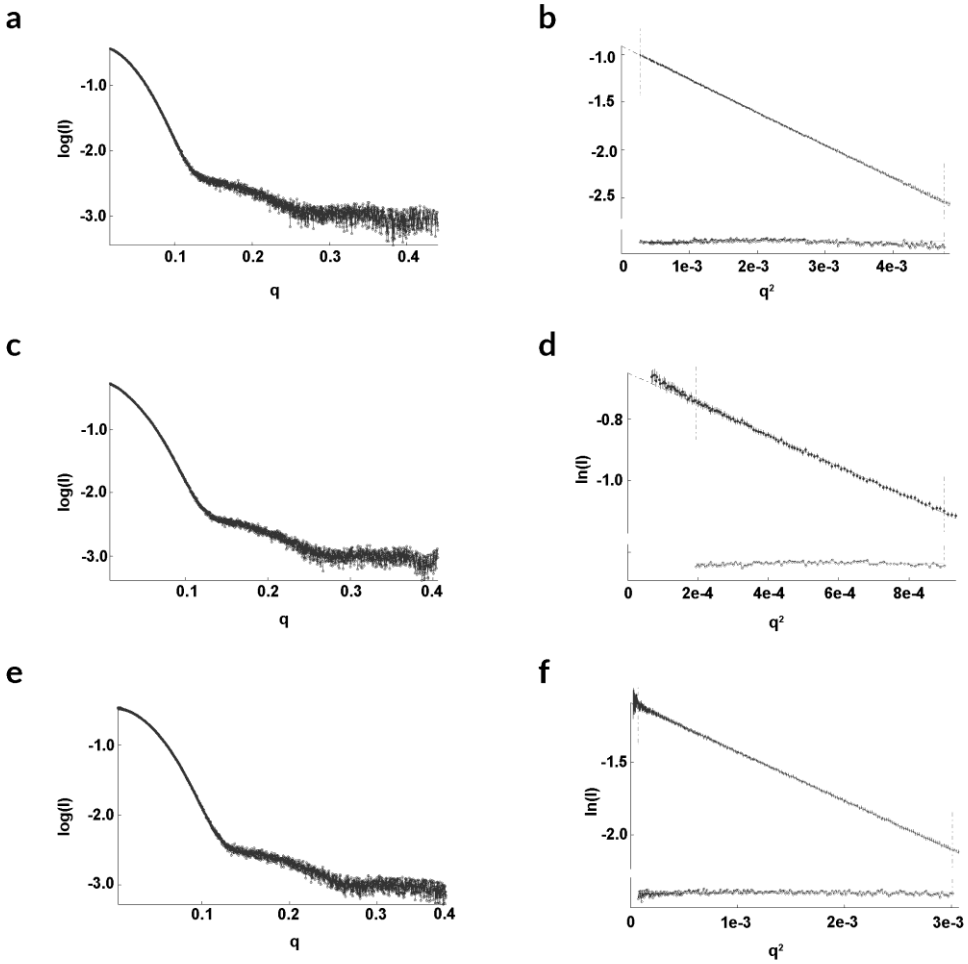


Figure 4.5, SAXS analysis for NEP at varied pH's. 1-dimensional SAXS profiles for NEP are shown for pH 5.5 (a), pH 7.5 (b) and pH 9.5 (e). Guinier plots for pH 5.5 (b), pH 7.5 (d) and pH 9.5 (f) were used to analyse the linear region of the low q data.

The effect of salt concentration on NEP conformation was analysed using HPLC mode SAXS. NEP was analysed at 2.5 mg/mL using high (500 mM NaCl) and low (150 mM NaCl) salt conditions. The size exclusion elution trace for high and low salt conditions is shown in Figure 4.6. A clear peak corresponding to NEP was observed in both size exclusion traces after 400 seconds. The first part of the protein peak in both purifications was not selected for analysis as signs of low-level aggregation were indicated with an increase to R_g value. Collected SAXS datasets were then plotted onto a 1-dimensional profile (Figure 4.7a and c) before using a Guinier analysis to estimate R_g values (Figure 4.7b and d). Again, R_g and calculated D_{max} values are shown in Table 4.2.

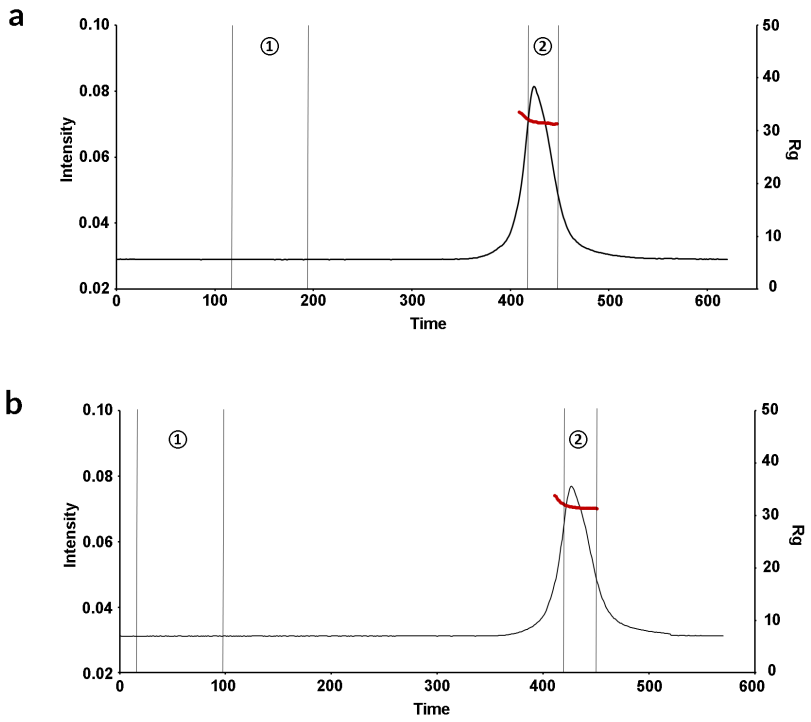


Figure 4.6, NEP SEX-SAXS chromatograms. (a) Elution profile for 150mM and (b) 500mM NEP. The region of the elution peak analysed is maked by ②, and the area used for buffer substration ①. A red line is used to indicate the estimated R_g of protein across the elution peak.

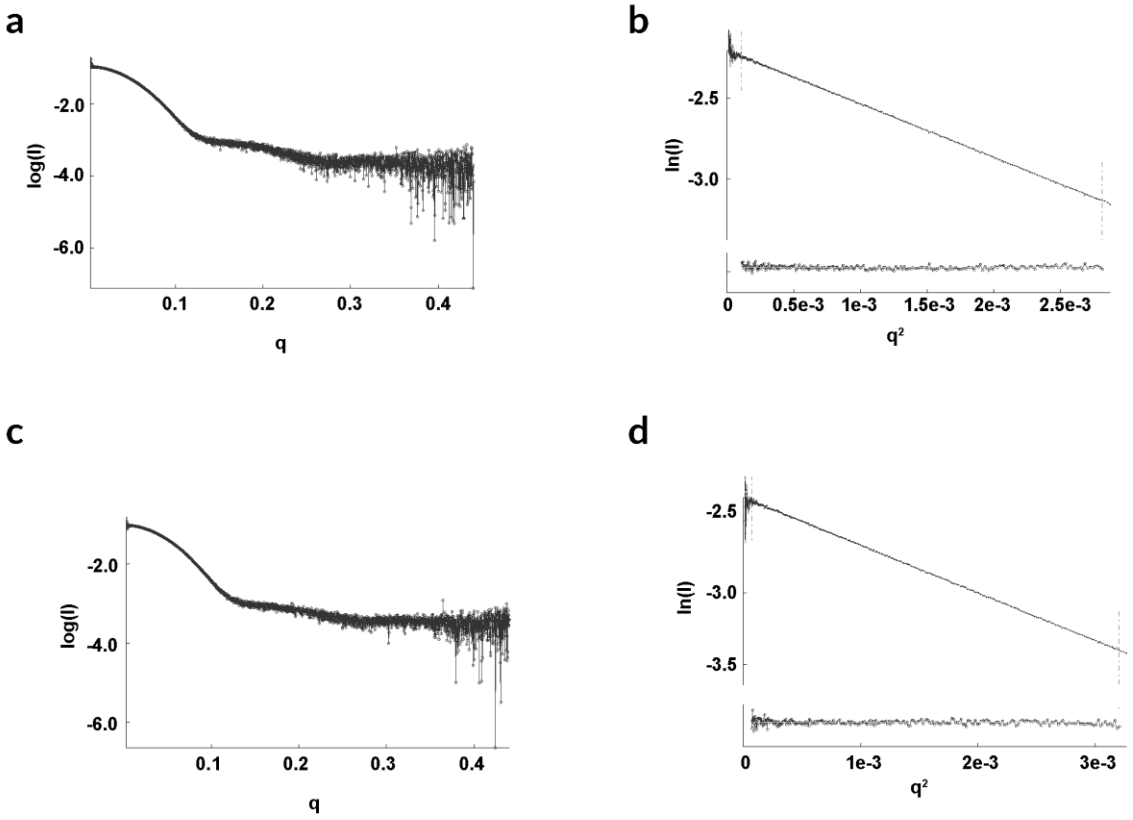


Figure 4.7, SAXS analysis for NEP at varied salt concentrations. 1-dimensional SAXS profiles are shown for 150mM NaCl (a) and 500mM NaCl (c). Guinier plots for 150mM NaCl (b) and 500mM NaCl (d) were used to analyse the linear region of the low q data.

4.4.2 Analysis of scattering parameters

Scattering parameters calculated from NEP SAXS data are shown in Table 4.2. In addition to the experimentally calculated scattering parameters, predicted scattering parameters were also calculated using the crystal structure of substrate-free NEP (6GID) and SAXS software Crysol. Using these scattering parameters, it was possible to analyse the correlation between experimental and predicted data.

	I(0)	***R_g***	***D_{max}***
Batch mode			
NEP pH 5.5	0.52252	39.1	189
NEP pH 7.5	0.40194	32.3	101.98
NEP pH 9.5	0.33505	31.8	103.76
HPLC mode			
NEP 150mM NaCl	0.11035	31.5	107.44
NEP 500mM NaCl	0.09633	31.5	107.28
Predicted			
6GID	-	28.42	105.7

Table 4.2, Calculated SAXS scattering parameters for NEP. Estimated R_g and D_{max} values are given in Å.

Figure 4.8 shows a Kratky plot of the experimental batch mode data collected for NEP at pH 7.5 and the predicted data from the crystal structure of substrate-free NEP. A clear correlation can be seen between these datasets with a X^2 value of 1.154 for $q < 0.4$. Particular focus should be placed onto the low q data ($q < 0.15$) which provides information regarding the size and shape of the protein particle. This data has a substantially lower level of noise and has a X^2 value of 2.503 indicating a similar particle shape and size for both experimentally calculated and predicted data.

The correlation between predicted and experimentally calculated data results in calculated scattering parameters which were similar with R_g values of 28.42 and 32.3 and D_{max} values of 105.7 and 101.98 respectively (Table 4.2). While these parameters are similar there is some discrepancy particularly in the R_g value. Several factors may be responsible for this

discrepancy. First, the crystal structure of substrate-free NEP is does not include 19 N-terminal amino acid residues due to their flexibility. Because of the absence of information regarding these residues it was not possible to included them when calculating predicted scattering data for substrate-free NEP. These 19 residues would however be contributing to experimental scattering and likely account for some of the discrepancy in scattering parameters. Second, mass spectrometry data in chapter 3 revealed the glycosylation profile of NEP to be varied between molecules. While attempts were made to accurately model glycosylation onto the substrate-free NEP crystal structure, it is reasonable to assume that some error would be present in the predicted scattering data.

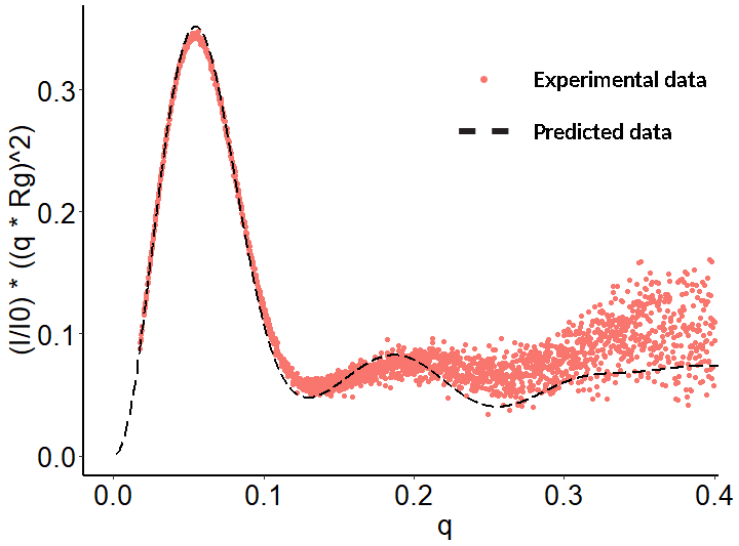


Figure 4.8, SAXS Kratky plot for NEP. SAXS batch mode data from NEP at pH 7.5 is plotted with the predicted normalised scattering data from the substrate-free NEP crystal structure (6GID).

A high degree of correlation was observed for scattering parameters calculated from SAXS data collected for NEP in varied buffer conditions. One exception to this was observed for the dataset collected at pH 5.5. This dataset had significantly higher calculated R_g and D_{max} values. To identify the cause of this variation, Dammif was used to visualise changes to the protein envelope (Figure 4.9). Calculated envelopes indicated that the increased scattering parameters were due to a long protrusion from the protein while the core of the protein particle remained unchanged. This was most likely due to pH dependent protein unfolding at the flexible N-terminus and not a result of subdomain separation for substrate access.

It should also be noted that D_{max} values appear slightly higher for SAXS data collected using HPLC mode. This may be a consequence of the size exclusion purification and rigid selection of a specific regions of the protein elution peak. The small variation may therefore correspond to scattering of differently glycosylated species.

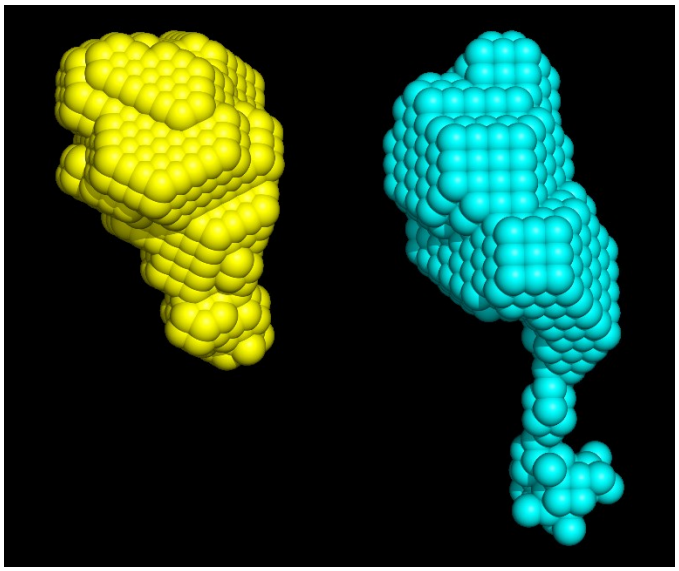


Figure 4.9, 3-dimensional NEP envelope models. Batch mode SAXS dataset collected for NEP at pH 7.5 (yellow) and pH 5.5 (cyan) are analysed and modelled using Dammif.

4.5 Conclusion

SAXS analysis was aimed to validate the crystal structures of NEP and to identify possible conformations that exist in solution. Validation of crystal structures was achieved through comparison of experimentally calculated and predicted SAXS data. The similar scattering parameters for this data suggests that in solution NEP adopts similar structure to that observed in the crystal structure. These findings also indicate that NEP does not exist in an open conformation with separate subdomains.

To identify whether detectable conformational changes in NEP could be provoked, variations in protein buffer conditions were investigated. Despite the use of these varied conditions, significant changes in scattering parameters were not observed with the exception of NEP analysed in a pH 5.5 buffer. The changes observed in this condition, however, where likely due to N-terminal unfolding and not a result of subdomain separation. NEP adopted a conformation in a variety of buffer solutions that resembled that

of the closed conformation observed in the crystal structure. This seemingly stable closed conformation raises questions regarding the ability of substrates to enter the central protein cavity for degradation.

Several hypotheses may explain the experimentally calculated SAXS data while addressing the necessity for conformational motion. First, subtle conformational movements may be present in NEP that are sufficient to allow for substrate entry but not large enough to result in significantly different scattering parameters. Second, a small proportion of protein may adopt an open conformation, but this conformation is masked by the majority of protein that exists in a closed conformation. Third, substrate interaction may be required for conformational movement and access to the active site.

To address these questions several avenues could be explored. More subtle conformational movements could be investigated through the analysis of deglycosylated NEP protein. A deglycosylated protein should provide more precise scattering parameters for NEP and remove any issues of mixed glycosylation species. In a similar way, collecting SAXS data for an NEP protein containing just the amino acids modelled in the crystal structure (19 N-terminal amino acid residues truncated) would remove inaccuracies when comparing experimentally calculated and predicted SAXS data. Likewise, the requirement of substrate interaction for conformational motion could be investigated through the addition of a peptide substrate to protein buffer during a SAXS experiment.

Ultimately, however, it may be necessary to investigate the conformational changes using cryo-EM. Cryo-EM may allow determination of the exact conformational movement in NEP and reveal conformations of NEP that may be obscured through the averaging nature of SAXS scattering data.

Chapter 5 – Characterisation of peptide-binding interactions in neprilysin

5.1 Foreword

This chapter is a continuation of the work included in Chapter 3 and describes the methods used to determine a peptide-bound structure of neprilysin. The resulting peptide-bound crystal structure is included as a published manuscript - *Moss, S., Subramanian, V. and Acharya, K. R. (2019), Crystal structure of peptide-bound neprilysin reveals key binding interactions. FEBS Lett. 594, 327-336*. Like Chapter 3, more extensive background work has been included in addition to a more detailed introduction with the aim of providing greater context to the work within this thesis.

5.2 Introduction

To date, over 30 peptide substrates have been identified for neprilysin (NEP) many of which link the protein to cardiology and neurology (Table 5.1) (Erdös and Skidgel, 1989). Two classes of substrate in particular, amyloid-β and natriuretic peptides (ANP, BNP and CNP), have led to considerable interest in NEP, establishing it as a possible therapeutic target for Alzheimer's disease and cardiovascular disease respectively (El-Amouri *et al.*, 2008; Jhund and McMurray, 2016).

Vasoactive peptides	**Neurological peptides**	**Other peptides**
Adrenomedullin	Amyloid-β	Caspase-9 precursor
Angiotensin 1 & 2	Cholecystokinin	Gastrin
Bradykinin	Corticotropin	Glucagon
Endothelin 1, 2 & 3	Dynorphin	
Kallidin	Galanin	
Natriuretic peptides (ANP, BNP, CNP)	Oxytocin	
Neurokinin	Melanin-concentrating hormone	
Neuropeptide	Melanotropin	
Physalaemin	Neurotensin	
Substance P	Enkephalin	
	Peptide YY	

Table 5.1, Peptide substrates of NEP with known vasoactive or neurological functions. Substrate list based on the MEROPS database (Rawlings et al., 2018) and Campbell (2017).

The identified substrates of NEP display large amounts of variation in charge, shape and sequence. Despite this variation, all substrates are uniform in their small size (< 5 kDa). This size specificity shown by NEP is strongly linked to the size of its central cavity (Figure 5.1). It is the limited volume of this cavity that permits only small peptides entry and subsequent cleavage (Macours *et al.,* 2004). In addition to cavity size, the electrostatic potential of the cavity likely plays a role in substrate specificity. Indeed, electrostatic interactions are known to be a key mechanism for many binding interactions and long range attraction (Plantinga *et al.,* 2008; Zhou and Pang, 2018).

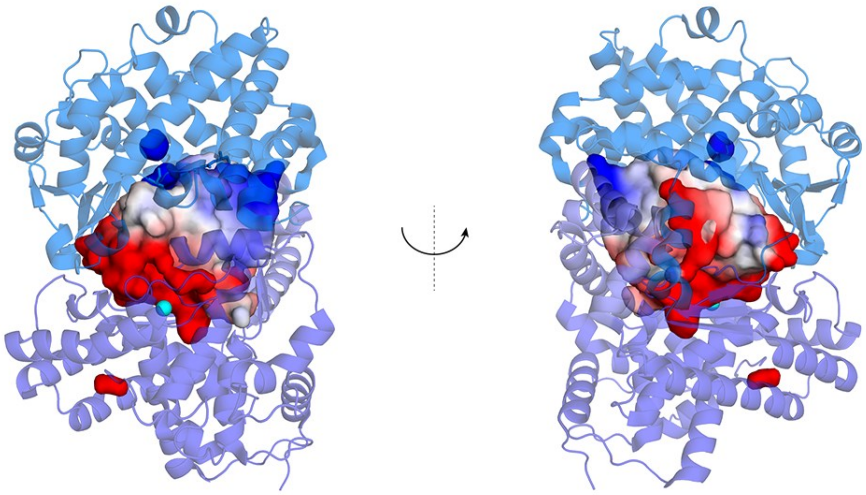


Figure 5.1, Crystal structure of substrate-free NEP (6GID). NEP contains a large central cavity with a distinct electrostatic potential. The active site of NEP is located within the cavity and is based around a zinc ion (cyan).

While substrates are clearly diverse in nature and the mechanism for substrate specificity is not fully understood, research has revealed that NEP exhibits a preference in cleavage positions shared across substrates (Quay *et al.,* 1994; Spungin-Bialik *et al.,* 1996). That is, NEP commonly cleaves peptides at the amino side of hydrophobic residues. Figure 5.2 is a schematic representation of the NEP subsites and substrate residue positions according to the Schechter and Berger convention (Schechter and Berger, 1967). In the schematic the S1′ site has the preference for a hydrophobic residue as such the P1′ residue is commonly a phenylalanine, leucine, tyrosine, isoleucine or valine (Rawlings *et al.,* 2018). In addition to the S1′, there is also some degree of specificity contributed by S1 site which has presence for a glycine residue in the P1 position. It is only these two subsites that have been

determined to have a substantial effect on cleavage position. The identified cleavage sites for well-known substrates - natriuretic peptides and amyloid-β - largely obey these recognised preferences (Figure 5.3).

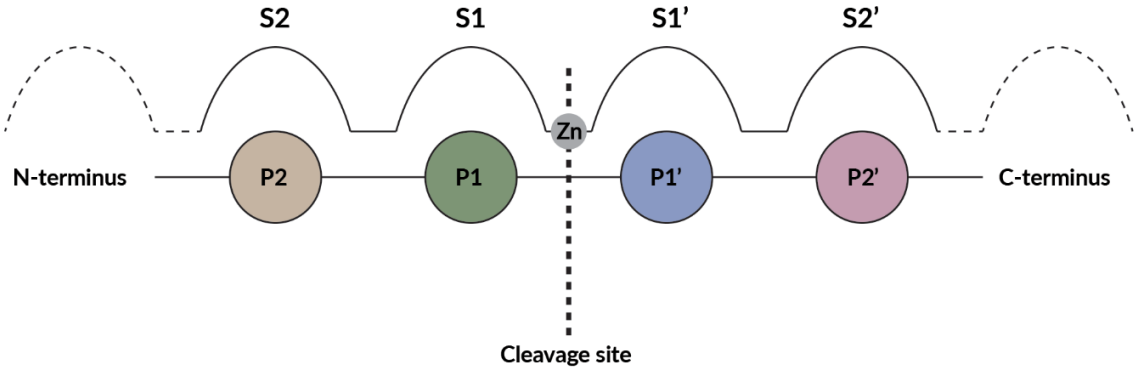


Figure 5.2, Schematic of NEP subsites and corresponding substrate residues. Subsites (S) and substrate residue positions (P) are labelled based on cleavage position. Numbers are non-prime in the direction of the substrate N-terminus and prime (') in the direction of C-terminus. The labelling convention follows the Schechter and Berger convention (Schechter and Berger, 1967)

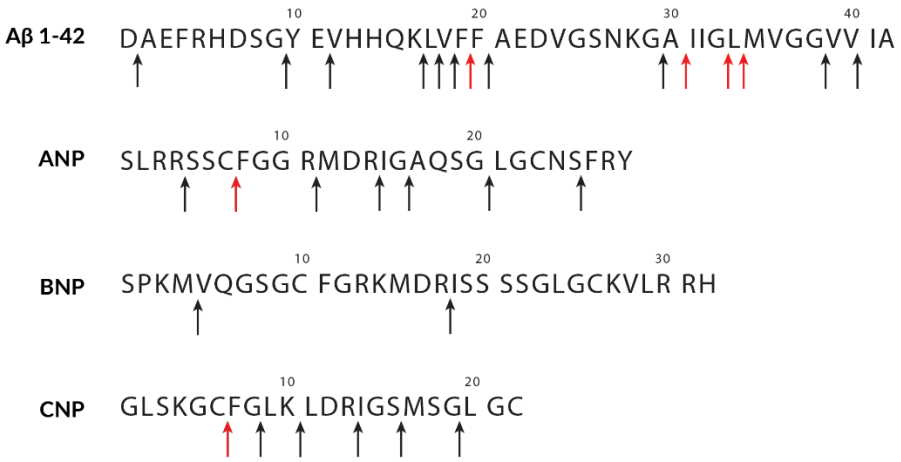


Figure 5.3, Location of NEP cleavage sites in amyloid-β and natriuretic peptides. Cleavage site locations in amyloid-β (Rogeberg *et al.,* 2014) *natriuretic peptides (atrial, B-type and C-type)* (Potter, 2011) *are indicated by arrows. Red arrows are used to indicate the proposed primary cleavage locations.*

Crystal structures for NEP have been published in a substrate-free and inhibitor-bound state. The peptide like design of inhibitors including phosphoramidon (Figure 5.4) has made

it possible to characterise subsites S1-S2′ and identify residues that are likely to be involved in substrate binding (Figure 5.5). However, the absence of a substrate-bound structure prevents complete understanding of the substrate binding in NEP.

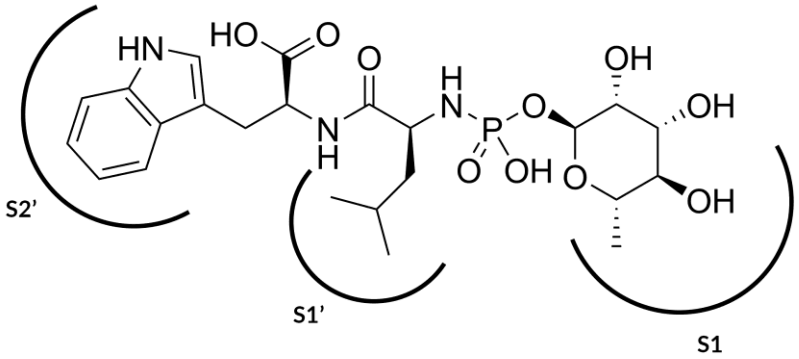


Figure 5.4, Zinc metalloprotease inhibitor phosphoramidon. The inhibitor is peptide like in design and the substrate scissile bound is mimicked by the bound between the phosphorus and nitrogen. Sections of the inhibitor are labelled with the subsites in NEP that form interactions.

A substrate-bound structure would allow several of the underlying questions regarding substrate binding in NEP to be addressed. Firstly, the general mode of peptide binding could be identified. That is, the position of the peptide backbone and the interactions to NEP amino acids. This general mode of binding is likely to be highly similar for all substrates of NEP, a feature that has been demonstrated in other zinc metalloproteases (Shen *et al.*, 2006; Larmuth *et al.*, 2014).

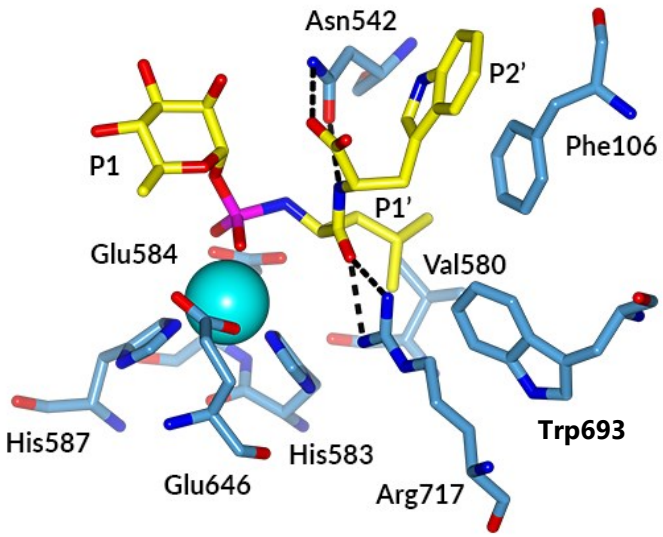


Figure 5.5, NEP active site with bound phosphoramidon. NEP active site residues (blue) and zinc ion (cyan) are shown with phosphoramidon (yellow). Phosphoramidon mimics a substrate and the corresponding substrate positions are labelled according to the Schechter and Berger convention.

The position of the peptide backbone would also confirm the orientation of peptides within the active site. While inhibitor binding strongly suggests cleavage occurs in the N- and C-terminal orientation shown in Figure 5.2, it is possible that cleavage may occur with the peptide flipped in the opposite direction. Some evidence for this, based on mass spectrometry experiments, is present in the zinc metalloprotease angiotensin-converting enzyme (Larmuth *et al.,* 2014).

Secondly, a substrate bound structure would allow for comparative investigation into the possible changes in peptide binding interactions that are generated from variation in substrate side chain residues. These could, in turn, be related to the observed preferences for residues in the P1 and P1' positions.

Thirdly, it would be possible to identify to what extent interactions between substrates and the NEP cavity occur in locations proximal to the active site. As current inhibitors occupy just the S1-S2' sites, a substrate-bound structure would reveal if there are binding interactions proximal to these sites which may have critical roles in substrate binding or positioning. While there is no experimental evidence to suggest proximal interactions are required, computation docking for NEP with natriuretic peptides has been conducted (Pankow *et al.,* 2009) and suggests that proximal residues including R222 and D377 may play a role in binding the peptide (Figure 5.6). It is equally possible that the majority of peptide binding interactions are contributed by the active site alone and that portions of substrate not directly bound to the active site have limited interactions with the cavity wall thus exhibiting a large amount of flexibility.

Finally, it has also been suggested that larger NEP substrates may not be entirely enclosed by the cavity and some degree of protrusion may be present. Structures in complex with long substrates may provide answers to this hypothesis.

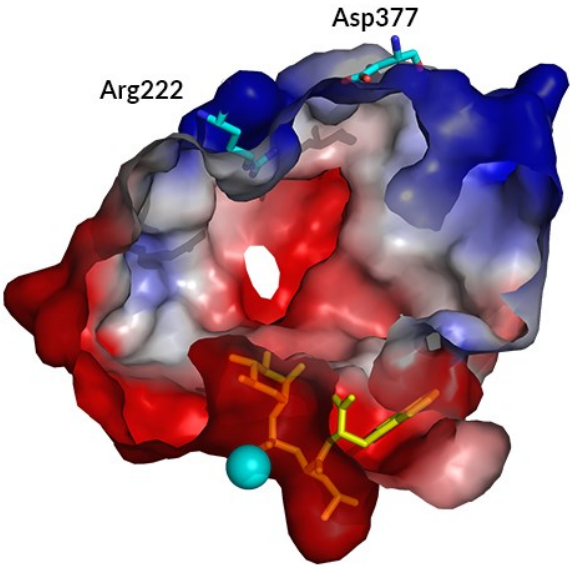


Figure 5.6, NEP cavity surface with phosphoramidon bound. Phosphoramidon interacts (yellow) with a shallow binding pocket formed of subsites S1-S2′ located around the zinc ion (cyan). In a substrate-bound structure a larger amount of the NEP cavity would be occupied. The interactions that take place in this cavity between substrate and NEP are unknown but docking studies have suggested Arg222 and Asp377 may have a role in substrate interaction (Pankow *et al.,* 2009).

5.3 Materials and Methods

All reagents for expression and purification were sourced from Sigma-Aldrich or Fisher Scientific unless otherwise specified. General methods were followed as described in Chapter 3 (3.2.1).

5.3.1 Plasmid construction and transfection

cDNA encoding full length human NEP in cloning vector pMD18-T Simple was purchased from Sinobiology. A polymerase chain reaction (PCR) was used to, amplify the extracellular domain of NEP (Thr49 - Trp749), add a 6x histidine tag and add restriction enzyme sites KpnI (5') and NotI (3'). The PCR reaction was set up in accordance with the NEB Q5 hot start polymerase manual. Specific component and parameter details are given in Table 5.2 and Table 5.3.

Component	Reaction Volumes	Final Concentration
5X Q5 Reaction Buffer	10 µL	1X
10 mM dNTPs	1 µL	200 µM
10 µM Forward Primer	2.5 µL	0.5 µM
10 µM Reverse Primer	2.5 µL	0.5 µM
1 ng/µL Template DNA	1 µL	0.02 ng/µL
Q5 Hot Start HF DNA Polymerase	0.5 µL	0.02 U/µL
Nuclease-Free Water	to 50 µL	

Table 5.2, PCR reaction components used to generate the histidine tagged NEP (Thr49 - Trp749).

Stage	Temperature	Time
Initial Denaturation	98 °C	30 seconds
35 Cycles ↓	98 °C	10 seconds
	58 °C	20 seconds
	72 °C	90 seconds
Final Extension	72 °C	2 minutes
Hold	4 °C	

Table 5.3, PCR reaction parameters used to generate the histidine tagged NEP (Thr49 - Trp749).

The amplified PCR product and vector pPICzαA were digested with restriction enzymes KpnI and NotI (NEP) in accordance with the NEB restriction enzyme protocol (Table 5.4). The PCR

product was then ligated into pPICzαA downstream of a secretion tag with T4 ligase (NEB) (Table 5.5).

Component	Reaction Volumes
Restriction Enzyme	1 µL (10U) per enzyme
DNA	1 µg
10X NEBuffer	5 µL (1X)
Nuclease-free water	to 50 µL
Incubation Time	1 hour
Incubation Temperature	37 °C

Table 5.4, Standard reaction parameters for a restriction digest.

Component	Reaction Volumes
T4 DNA Ligase Buffer (10X)	2 µL
Vector DNA (3 kb)	50 ng
Insert DNA (2.2 kb)	110 ng
T4 DNA Ligase	1 µL
Nuclease-free water	to 20 µL
Incubation Time	16 hours
Incubation Temperature	16 °C

Table 5.5, Ligation reaction parameters used to generate histidine tagged NEP pPICzαA.

Mutation of NEP residue 584 from glutamate to aspartate or glutamine (E584Q or E584D) was performed using PCR site-directed mutagenesis (NEB Q5). Specific component and parameter details are given in Table 5.6 and Table 5.7. Mutagenesis was confirmed using DNA sequencing. A full list of primer sequences used to generate the NEP E584D and E584Q is given in Table 5.8.

Component	Reaction Volumes	Final Concentration
Q5 Hot Start High-Fidelity 2X Master Mix	12.5 µL	1X
10 µM Forward Primer	1.25 µL	0.5 µM
10 µM Reverse Primer	1.25 µL	0.5 µM
1 ng/µL Template DNA	1 µL	0.02 ng/µL

Nuclease-Free Water	9 µL

Table 5.6, PCR reaction components used to generate NEP E584D and E584Q pPICzαA.

Stage	**Temperature**	**Time**
Initial Denaturation	98 °C	30 seconds
25 Cycles ↓	98 °C	10 seconds
	63 °C	20 seconds
	72 °C	90 seconds
Final Extension	72 °C	2 minutes
Hold	4 °C	

Table 5.7, PCR reaction parameters used to generate NEP E584D and E584Q pPICzαA.

Primer	**DNA sequence**
Sequencing primer 1	GCCAGGGTTTTCCCAGTCACGAC
Sequencing primer 2	CAGCTGAAAAAGCTATTGCACAACTGAATTC
Sequencing primer 3	GAGCGGATAACAATTTCACACAGG
Amplification Fwd (5′ KpnI and 6xHis)	CGCTGGTACCCATCACCACCACCATCACGAATTC
Amplification Rev (3′ NotI)	CGGACCTACGATGATGGTATTTGCAAG GCTTGCGGCCGCTCACCAAACCCGGCACTTC
E584Q mutagenesis Fwd	GACTAGGACACGA**C**ATCACCCATGGC
E584Q mutagenesis Rev	TGACCATGCCGATGCCCC
E584D mutagenesis Fwd	CAACATAGGACAC**C**AAATCACCCATGGCTTCG
E584D mutagenesis Rev	ACCATGCCGATGCCCCCA

Table 5.8, Primer DNA sequences for generation of NEP mutants. All primers used to generate and confirm the DNA sequence of NEP E584D pPICzαA and NEP E584Q pPICzαA are listed.

The mutant NEP pPICzαA constructs were transformed into TOP10 Chemically Competent *E. coli* cells (ThermoFisher) for plasmid replication. Transformation was performed by incubating 20 µL TOP10 cells on ice with 1 µL of 1 ng/µL NEP E584D pPICzαA or NEP E584Q pPICzαA DNA for 10 minutes. Cells were transferred to a water bath at 42°C for 45 seconds before incubating on ice for a further two minutes. 100 µL of LB was added to the cells and cells were incubated at 37°C for 45 minutes at 200 rpm. 100 µL cell solution was then spread

onto an LB agar plate containing 100 µg/mL ampicillin. LB agar plates were incubated overnight at 37°C.

5 mL of LB ampicillin (100 µg/mL) media was inoculated with transformed TOP10 colonies. 5 mL cultures were grown overnight at 37°C and 200rpm. Cultures were harvested by centrifugation at 3000 RCF for 10 minutes. DNA was harvested from TOP10 cells using Wizard® Plus SV Minipreps DNA Purification Systems (Promega).

NEP E584D pPICzαA or NEP E584Q pPICzαA DNA was linearised with restriction enzyme SacI (NEB) using the protocol described in Table 5.4. DNA was then integrated into the genome of *Pichia pastoris* GS115 using electroporation in accordance with the Invitrogen EasySelect manual (cat. no. K1740-01). Transformed cells were then plated onto a YPD plate containing 100 µg/mL zeocin for colony selection.

5.3.2 Protein expression and purification

Expression of recombinant NEP and NEP E584D extracellular domains from the *P. pastoris* GS115 was performed in accordance with the Invitrogen EasySelect manual. The protocol used was identical to that described in Chapter 3. For small-scale test expressions the standard protocol was used but the culture volume and induction time were scaled down to 25 mL and 24-hours.

5.3.3 Metal removal and replacement

Metal removal was attempted using chelating agents EDTA and 1,10 phenanthroline. NEP at 50 nM was buffer equilibrated into 25 mM Trizma base, 150 mM NaCl and 75mM EDTA or 1,10 phenanthroline (pH 7.5, pH adjusted with HCl) using a centrifuge filter unit with a 30kDa molecule weight cut-off (Millipore). NEP was incubated for 24-hours at 19°C before buffer equilibration into binding buffer (25 mM Trizma base, 150 mM NaCl, 2 mM $MgCl_2$ (pH 7.5, pH adjusted with HCl)).

For metal replacement NEP at 50 nM was buffer equilibrated into 25 mM Trizma base, 150 mM NaCl and 2 mM $CoCl_2$ or $NiCl_2$ (pH 7.5, pH adjusted with HCl). NEP was incubated for 24-hours at 19°C before buffer equilibration into binding buffer.

5.3.4 Enzyme assay

The activity of NEP and NEP E584D mutant were assessed by measuring cleavage of the fluorogenic peptide Mca-RPPGFSAFK-(Dnp) (Enzo Life Sciences). Protein was diluted to varied concentrations in binding buffer. 50 µL of diluted protein was added to a black 96 well microplate (cat. no. 655076, Greiner Bio-One). 50 µL of between 50 – 3.125 µM Mca-RPPGFSAFK-(Dnp), dissolved in binding buffer with 10% (v/v) DMSO, was added to a black 96 well microplate (Greiner Bio-One) to initiate the reaction.

Each fluorescence assay was performed in triplicate and fluorescence was recorded at 20 second intervals for at least 35 minutes on a CLARIOstar high-performance microplate reader (BMG LABTECK). Excitation was achieved at 328 nm and emission was detected at 393 nm.

Values for initial rates were determined in relative fluorescence units (RFU) using a line of best fit through data points up to 200 seconds. Where more than one substrate concentration was used, initial rate data was plotted, and a non-linear regression was used to estimate K_m and V_{max} values. RFU values were converted to µM using the conversion '1 µM = 2103 RFU' which was calculated using a standard curve of product fluorescence versus product concentration.

5.3.5 Circular dichroism

The secondary structure composition of NEP and NEP E584D mutant was analysed using circular dichroism. Protein samples of NEP and NEP E584D mutant were diluted to 2.725 µM and 2.6 µM respectively before analysis. Data were collected in triplicate using a 1 mm quartz cuvette and a Chirascan circular dichroism spectrometer at 20ºC. The wavelength scanning range was between 200-300 nm with a 1 nm step size, monochromator bandwidth of 3 nm and sample time-per-point of 3 seconds.

5.3.6 Crystallisation

Sacubitril-bound wild-type NEP - Protein crystallisation was performed by adding 1 µL of 50 mM sacubitril dissolved in 100% DMSO to 300 µL NEP at 0.75 mg/mL in binding buffer. The solution was incubated for 1 hour at 19°C before concentrating it to 30 µL using a centrifuge filter unit with a 30kDa molecule weight cut off (Millipore). The concentrated protein solution was mixed 1:1 with crystallisation buffer 0.2 M KNO_3 and 20% (w/v)

PEG3350 (Molecular Dimensions, England). 2 µL of the mixed solution was manually dispensed into 24-well (Molecular Dimensions, England) plates for crystallisation experiments using hanging drop vapour diffusion. Plates were stored at 18 ºC and crystals formed within a month. Crystals used for data collection were mounted in a loop and flash-cooled for storage in liquid nitrogen without cryoprotectant.

Peptide-bound wild-type NEP and NEP E584D - Protein crystallisation was performed by adding 4 µL NEP at 8-10 mg/mL to 1 µL of substrate dissolved in binding buffer. Substrates were Aβ fragments 1-40, 10-16, 35-42, 1-16, bradykinin (GenScript) and Mca-RPPGFSAFK-(Dnp) (Enzo Life Sciences). Substrate concentrations were varied to be between 6 – 34-fold excess of NEP in the final solution.

The protein solution was mixed 1:1 with crystallisation buffer 0.2 M KNO_3 and 20% (w/v) PEG3350 (Molecular Dimensions, England). The mixed solution was manually dispensed into 24-well (Molecular Dimensions, England) or 96-well (Intelli-Plate Art Robbins Instruments) plates for crystallisation experiments using hanging or sitting drop vapour diffusion. Plates were stored at 18 ºC and crystals formed within a month. Crystals used for data collection were mounted in a loop and flash-cooled for storage in liquid nitrogen without cryoprotectant.

Published peptide-bound NEP E584D CNP structure - Protein crystallisation was performed using sitting drop vapour diffusion. NEP E584D (8 mg/mL) was dispensed with a Phoenix crystallisation robot (Art Robbins Instruments) into an Intelli-Plate 96 (cat. no. 102-0001-03, Art Robbins Instruments) at a ratio of 0.2 µL: 0.2 µL protein to crystallisation buffer. Crystallisation buffer for the peptide-free structure was 0.16 M potassium nitrate, 0.04M sodium bromide, 0.02 M bis-tris propane and 20% (w/v) PEG 3350. Crystallisation plates were incubated at 18 °C and crystals formed within two months.

Attempts to co-crystallize NEP E584D with peptide were unsuccessful. Instead the peptide-bound structure was achieved through soaking. Crystals for soaking were generated using crystallisation buffer 0.16 M KNO_3, 0.04 M NaI and 20% (w/v) PEG 3350. Again, crystallisation plates were incubated at 18 °C and crystals formed within two months. Once formed 0.2 µl of 9.1 mM C-type natriuretic peptide was added to the crystallisation drops and incubated for a further 14 days at 18 °C before data collection. The C-type natriuretic peptide used had

the amino acid sequence 'GLSKGCFGLKLDRIGSMSGLGC' and was sourced from (GenScript). All crystals used for data collection were flash-cooled in liquid nitrogen for storage without the use of cryoprotectant.

5.3.7 X-ray data collection and refinement

X-ray diffraction data collection for the NEP E584D substrate-free and peptide-bound structures was performed on Diamond Light Source (Didcot, Oxford UK) beamlines I04 (mx17212-37) and I04 (mx17212-46) respectively. At the time of collection both beam lines were fitted with Eiger2 X 16M detectors.

X-ray diffraction data were indexed and integrated using DIALS (Waterman *et al.,* 2016). Data were scaled and merged using AIMLESS (Evans and Murshudov, 2013; Winn *et al.,* 2011) where resolution cut offs were applied. The structures were solved by molecular replacement using Phaser (McCoy *et al.,* 2007) and substrate-free neprilysin (6GID) (Moss *et al.,* 2018). Programmes Refmac5 (Murshudov *et al.,* 2011) and Coot (Emsley *et al.,* 2010) were used for refinement of the structures.

Validation of the structures was achieved using programmes MolProbity (Chen *et al.,* 2010) and PDB validation (Berman *et al.,* 2003). All structural images and superpositions were generated in CCP4mg (McNicholas *et al.,* 2011).

5.4 Results

5.4.1 Inhibitor-binding

Prior to substrate binding investigations, cocrystallisation of NEP was attempted with the inhibitor sacubitril. As a crystal structure was already available for NEP in complex with sacubitril, the cocrystallisation experiment was used to confirm that the previously determined crystallisation conditions (Chapter 3) were appropriate for cocrystal generation. It also provided an indication of the substrate concentration required for cocrystallisation.

Cocrystals of NEP and sacubitril were formed using the previously determined crystallisation conditions without any observable morphological differences to crystals formed in the absence of sacubitril. X-ray diffraction data for these crystals were collected to determine if inhibitor was bound. The highest resolution dataset was solved in space group $P3_221$ to a resolution of 2.5 Å. Full data collection and refinements statistics are shown in Table 5.9.

The determined structure displayed a high degree of similarity to the published substrate-free (6GID) and sacubitril-bound NEP (5JMY) with RMSD's for the 696 residues of 0.380 Å and 0.363 Å respectively. It was possible to model the sacubitril molecule into electron density proximal to the catalytic zinc ion (Figure 5.7a). A superposition of the published sacubitril-bound NEP structure (5JMY) and the structure presented here indicates the almost identical positions of inhibitor and active site residues (Figure 5.7b). The only notable difference in the two crystal structures was that 5JMY was crystallised in the orthorhombic space group $P2_12_12_1$ and not $P3_221$ as reported here.

As sacubitril binding in NEP has been reported in detail for the 5JMY structure (Schiering *et al.,* 2016), further analysis is not given here. However, the presented unpublished structure confirmed that the crystallisation conditions and crystallisation protocol were appropriate for generating ligand-bound NEP crystals. As such, it was possible to begin peptide cocrystallisation.

	Sacubitril-bound NEP
Beamline	I03, DLS
Wavelength (Å)	0.9763
Crystallographic statistics	
Space group	$P3_221$
Unit cell dimensions	
a,b,c (Å)	107.93, 107.93, 112.53
α, β, γ (°)	90.00, 90.00, 120.00
Resolution range (Å)	71.90-2.50 (2.60-2.50)
R merge	0.254 (5.417)
$CC_{1/2}$	0.999 (0.539)
⟨I/σ (I)⟩	11.4 (1.1)

Completeness (%)	100.0 (100.0)
No. observed reflections	1044509 (117321)
No. unique reflections	26709 (2989)
Multiplicity	39.1 (39.3)
Refinement statistics	
R_{work} /R_{free} (%)	0.240/0.277
R.m.s deviations	
Bond lengths (Å)	0.004
Bond angles (°)	1.183
Ramachandran statistics	
Favoured (%)	98
Allowed (%)	2
Outliers (%)	0
Wilson B-factor (Å^2)	39.9
Average B-factors (Å^2)	
Protein	75.36
Ligand	70.33
Water	61.32
Carbohydrate	86.73
No. of atoms	
Protein	11226
Ligand	28
Water	7
Carbohydrate	42

Values in parentheses are for highest-resolution shell.

Table 5.9, Data collection and refinement statistics for NEP in complex with sacubitril.

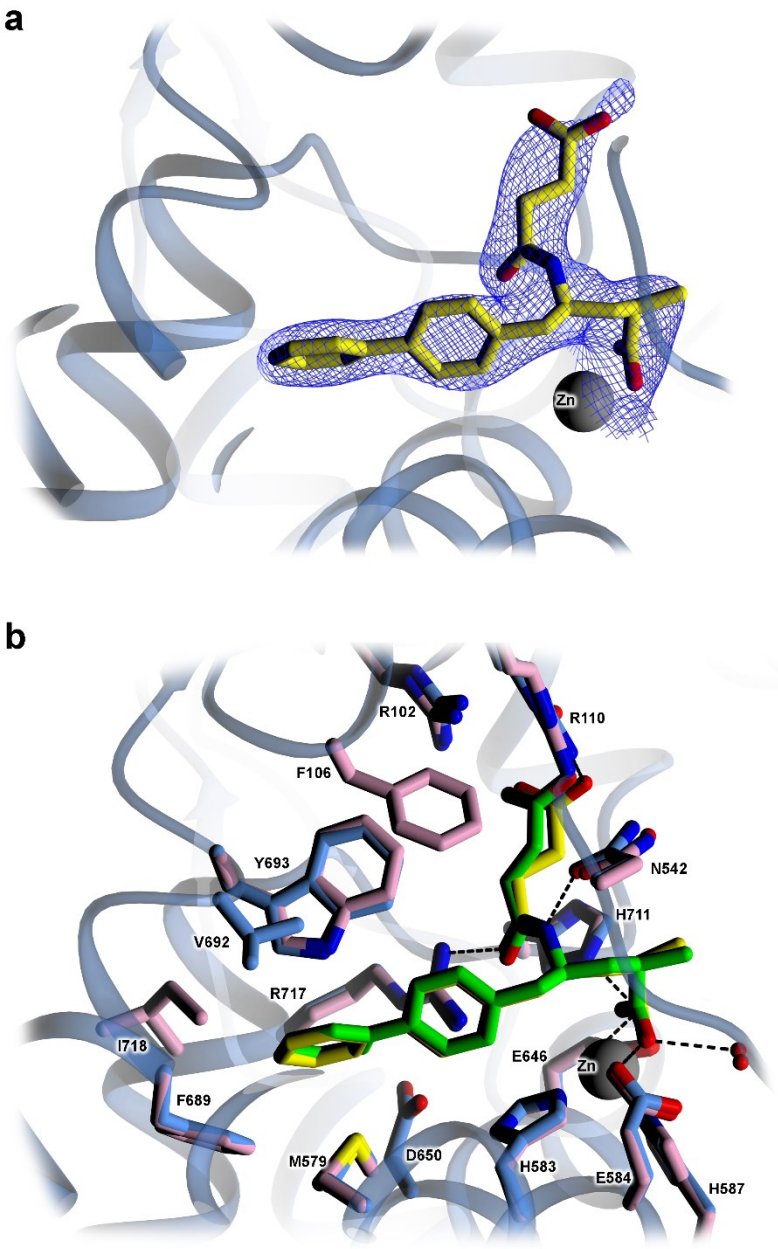


Figure 5.7, NEP active site with bound inhibitor sacubitril. (a) 2Fc-Fo electron density map (contoured to 1σ) supported the placement of sacubitril. (b) superposition of published sacubitril-bound NEP structure (5JMY, pink and green) and unpublished structural data presented in this chapter (blue and yellow).

5.4.2 Peptide-binding

Given that cocrystallisation of NEP with an inhibitor was possible and density for the bound inhibitor in the active site was clear, cocrystallisation with several substrates was performed. These substrates included bradykinin, fluorescent Mca-RPPGFSAFK-(Dnp) peptide and amyloid-β fragments 1-40, 10-16, 1-16 and 35-42. Different amyloid-β fragments were used due to the presence of multiple cleavage sites in the full-length peptide. As NEP can bind

and cleave amyloid-β at several locations, smaller fragments were used to limit the number of possible binding positions that could be observed in a crystal structure.

NEP cocrystals were grown in the presence of substrates and displayed the same crystal morphology as crystals grown in the absence of substrate. As described in Chapter 3, crystals exhibited variation in diffraction quality within different areas of the same crystal and between different crystals. Given this range in diffraction quality many cocrystals were screened with only a small number better than 3 Å in resolution. In general, lower resolution data was far less likely to have strong density corresponding to a bound ligand. While high-resolution datasets were more likely to have stronger density for a bound ligand, they frequently did not.

Due to the variation seen in resolution and peptide binding, a large number of datasets from NEP cocrystals were collected. These datasets were solved using molecular replacement and refined using a single refmac5 run. This fast processing protocol was used to quickly determine if structures contained unmodelled electron density corresponding to bound peptide. Four structures have been included to illustrate the commonly observed electron density proximal to the zinc ion in a position that would be occupied by bound peptide (Figure 5.8).

The majority of solved structures had a small amount of electron density above the zinc no larger than a water molecule (Figure 5.8a). Other structures contained slightly elongated electron density, but the density was ambiguous and 'not peptide-like' in shape (Figure 5.8b). In these cases, it is likely that different molecules were bound to the zinc across the NEP molecules that make up a crystal. The electron density is therefore an average of these different molecules gives rise to the ambiguous observed density.

The determined structures from two of the highest resolution datasets collected (1.9 Å and 2.1 Å) were identified to have large areas of unmodelled electron density (Figure 5.8a and b). It seems likely that higher resolution data is required to observed clear electron density corresponding to bound peptide. In these two higher resolution structures the electron density proximal to the zinc ion shared a common peptide-like shape. To better understand the identity of this electron density a superposition of the structure was made with phosphoramidon-bound NEP (1DMT) (Figure 5.9a). As phosphoramidon has a peptide-like

structure and the fact it overlapped with the observed electron density in the P1′ and P2′ positions supported the hypothesis that the density corresponded to a bound peptide.

Using the superimposed phosphoramidon structure as a template, an alanine dipeptide fragment was modelled into the electron density (Figure 5.9b). The modelled peptide provided a convincing description of the observed electron density, however the peptide could be positioned with the C- or N-terminus towards the zinc ion. Figure 5.9b shows the C-terminus orientation as the density fit was moderately better.

From the electron density it was clear that there was no continuation in peptide past the P1′ and P2′ positions. This observation suggested peptide cleavage had occurred and the bound peptide was likely to be a small dipeptide fragment. This observation was not unexpected given that the NEP protein used for crystallisation had been confirmed to possess protease activity. Peptide cleavage did however represent a major issue in observing peptide binding interactions. With this in mind the possibility to remove protease activity from NEP was explored.

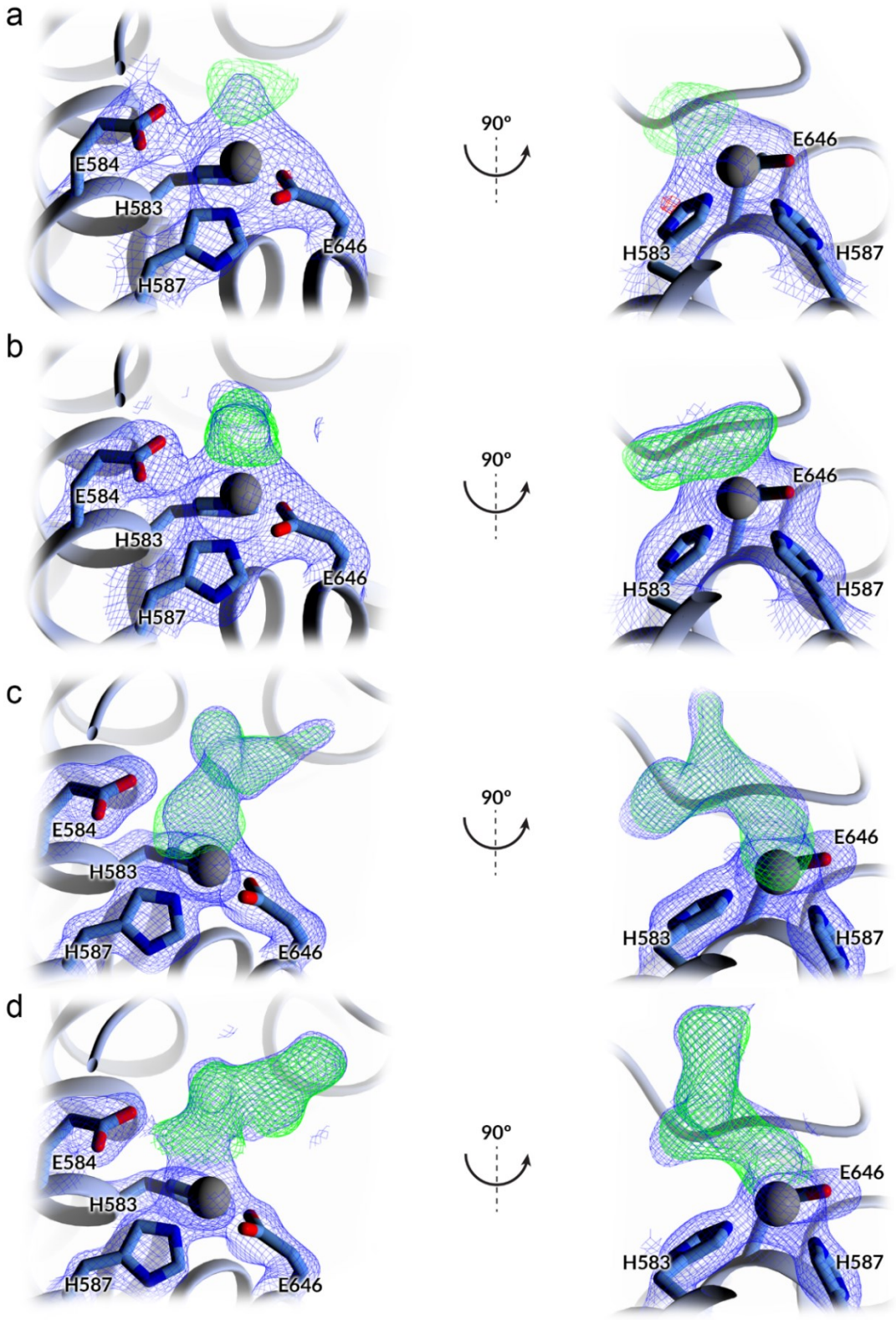


Figure 5.8, Unmodelled electron density in wild-type NEP active site. NEP crystals formed in the presence of peptide substrates. (a) amyloid-β fragment 35-42, (b) amyloid-β fragment 1-16, (c) bradykinin and (d) fluorescent Mca-RPPGFSAFK-(Dnp) peptide. 2Fc-Fo and Fc-Fo electron density maps are contoured at 1 and 3σ respectively.

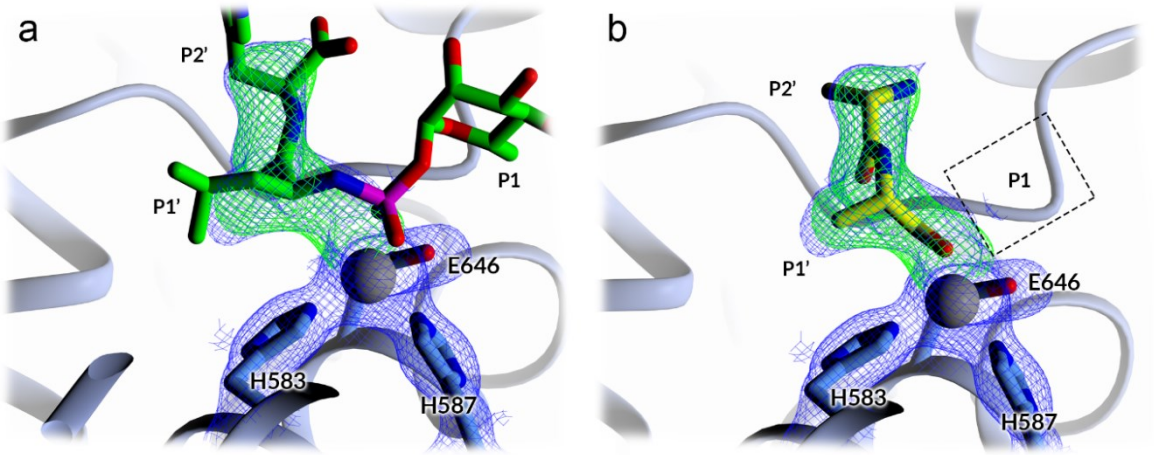


Figure 5.9, Modelled electron density in wild-type NEP active site. (a) Phosphoramidon-bound NEP structure (1DMT) is superimposed onto the determined structure and the position of the bound phosphoramidon is shown (green). (b) di-alanine peptide (yellow) is modelled into the electron density using the phosphoramidon coordinates as a template. The electron density correlates well to a bound peptide and it is clear that no continuation in the peptide chain is present past the cleavage site (dashed box). Fc-Fo and 2Fc-Fo electron density maps are contoured \t 1 and 3σ respectively.

5.4.3 NEP metal removal and replacement

The proposed mechanism for zinc metalloprotease enzymes is centred around the zinc ion and the removal of this ion has been shown to eliminate protease activity (Holmquist and Vallee, 1974). Zinc removal from metalloproteases has been performed using chelating agents including phenanthroline and EDTA (Bünning and Riordan, 1985). It has also been shown that the zinc ion can be removed through mutation of a zinc binding residue. Such a method has been used on the zinc metalloprotease insulin-degrading enzyme (IDE) where mutagenesis of a conserved histidine residue in the HEXXH motif was used to generate IDE protein absent of the zinc ion. While the mutation removed protease activity, the protein retained the secondary structure and insulin binding capacity (Gehm *et al.,* 1993). Subsequent X-ray crystallography research into IDE has resulted in structures of several substrate-bound structures of IDE absent of a zinc ion (Shen *et al.,* 2006). The success of metal removal in IDE indicated such a method would be a good target for use with NEP.

For NEP, metal removal was initially attempted using chelating agents and not mutagenesis. The major benefit of this being that the same wild-type protein could be used, and new mutant protein did not have to be expressed. For metal removal NEP protein was incubated with chelating agents phenanthroline or EDTA for 24 hours. These components were then

removed using buffer exchange. A cleavage assay, as described in Chapter 3, was then performed on NEP samples (Figure 5.10 and Figure 5.11). From the assay it was clear that treatment with chelating agents was unsuccessful in removing NEP protease activity. While some variation was seen in the initial rate of reaction, substantial peptide cleavage was still present after each treatment. It is unclear why protease activity was retained; it is possible that the zinc ion was very stably bound in NEP preventing its removal.

While metal removal was unsuccessful, the possibility to replace the zinc ion with an alternative metal was explored. Metal substitution has been shown to alter activity in other zinc metalloproteases (Holmquist and Vallee, 1974; Bünning and Riordan, 1985). Nickel and cobalt were selected for metal replacement in NEP. In model zinc metalloprotease thermolysin, cobalt substitution resulted in an increase in activity while nickel replacement resulted in a <2% activity. NEP zinc replacement was performed by incubating NEP with an excess of nickel and cobalt in order to outcompete zinc. Excess metal was then removed using buffer exchange. Like treatment with chelating agents, metal replacement also yielded NEP with retained protease activity (Figure 5.10). Again, some variation was seen in the initial rates (Figure 5.11) however, as protease activity was still present, neither metal removal nor replacement was appropriate for production of inactive NEP that could be used for substrate cocrystallisation.

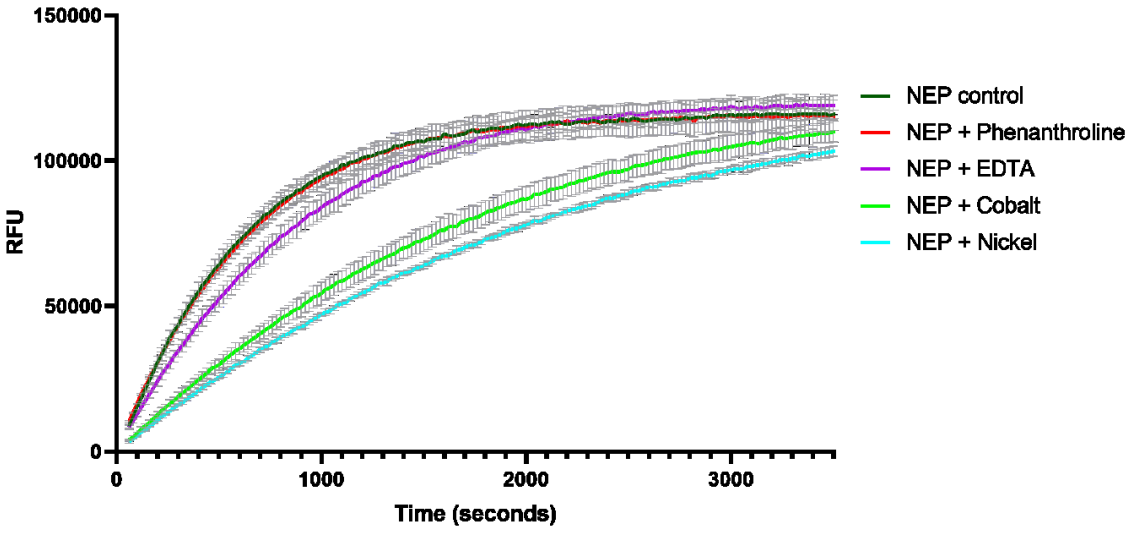


Figure 5.10, NEP peptide cleavage assay post zinc removal or replacement treatments. NEP was incubated with 75mM chelating agents phenanthroline or EDTA, or 2mM metals cobalt or nickel. The treated NEP at 3.125 nM (0.256 mg of protein in a 100 µL reaction volume) was then incubated with 50 µM peptide substrate Mca-

RPPGFSAFK-(Dnp) to assess peptide cleavage. Some variation was observed in cleavage between NEP samples however all samples still retained protease activity. Curves are plotted with the mean value and ±SD error bars for three repeats.

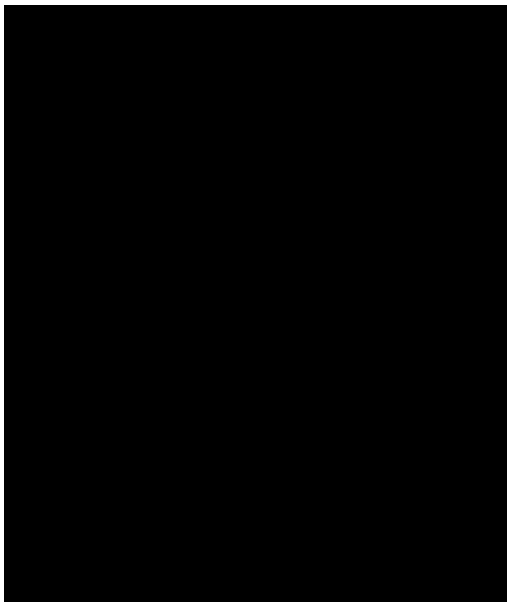


Figure 5.11, NEP peptide cleavage initial rates after zinc removal or replacement treatments. Initial rates were calculated from a single 50 µM substrate concentration. Mean values and ±SD error bars are plotted for three repeats.

5.4.4 NEP glutamate 584 mutagenesis

Removal of protease activity in zinc metalloproteases has also been demonstrated through mutagenesis. As previously stated, it is possible to mutate one of the conserved histidine residues in the HEXXH motif to prevent zinc binding and therefore remove protease activity. While this method has been shown to inactive zinc metalloproteases and prevent substrate cleavage, the absence of the zinc ion likely has the undesirable effect of altering substrate binding. For this reason, an alternative target for mutagenesis, the conserved glutamate in the HEXXH motif, was used. The glutamate is involved in hydroxide generation during proteolysis, but the residue has limited effect on substrate binding in other zinc metalloproteases.

Glutamate mutations E584D and E584Q were selected with the aim to conserving the amino acid characteristics. To generate mutant NEP E584D and E584Q it was necessary to first use molecular cloning to produce the appropriate DNA and then transform that DNA into *P*.

pastoris for protein expression. An overview of the cloning method is shown in Figure 5.12. Briefly, PCR was used to amplify the DNA sequence corresponding to extracellular domain from the full-length NEP gene. During amplification restriction sites KpnI (5′) and NotI (3′) were added in addition to a 6x histidine tag (5′) (Figure 5.13a). The resulting amplification product and vector pPICzαA were double digested with KpnI and NotI before being ligated together. The ligation product was transformed into TOP10 *E. coli* cells before growth on a zeocin LB plate to select colonies containing NEP pPICzαA. A single colony was used for DNA production and the harvested DNA was digested to confirm the presence of the NEP gene within the vector (Figure 5.13b).

The NEP pPICzαA construct was then targeted for site directed mutagenesis using PCR and converting glutamate 584 to an aspartate or glutamine. The resulting PCR products were sequenced to confirm successful point mutations had been achieved and that no other mutations had resulted from the PCR. The mutant NEP E584D pPICzαA and NEP E584Q pPICzαA constructs were then linearised using restriction enzyme SacI before transformation and DNA integration into *P. pastoris* using electrophoresis. Transformed *P. pastoris* cells were then plated onto YPD zeocin containing plates.

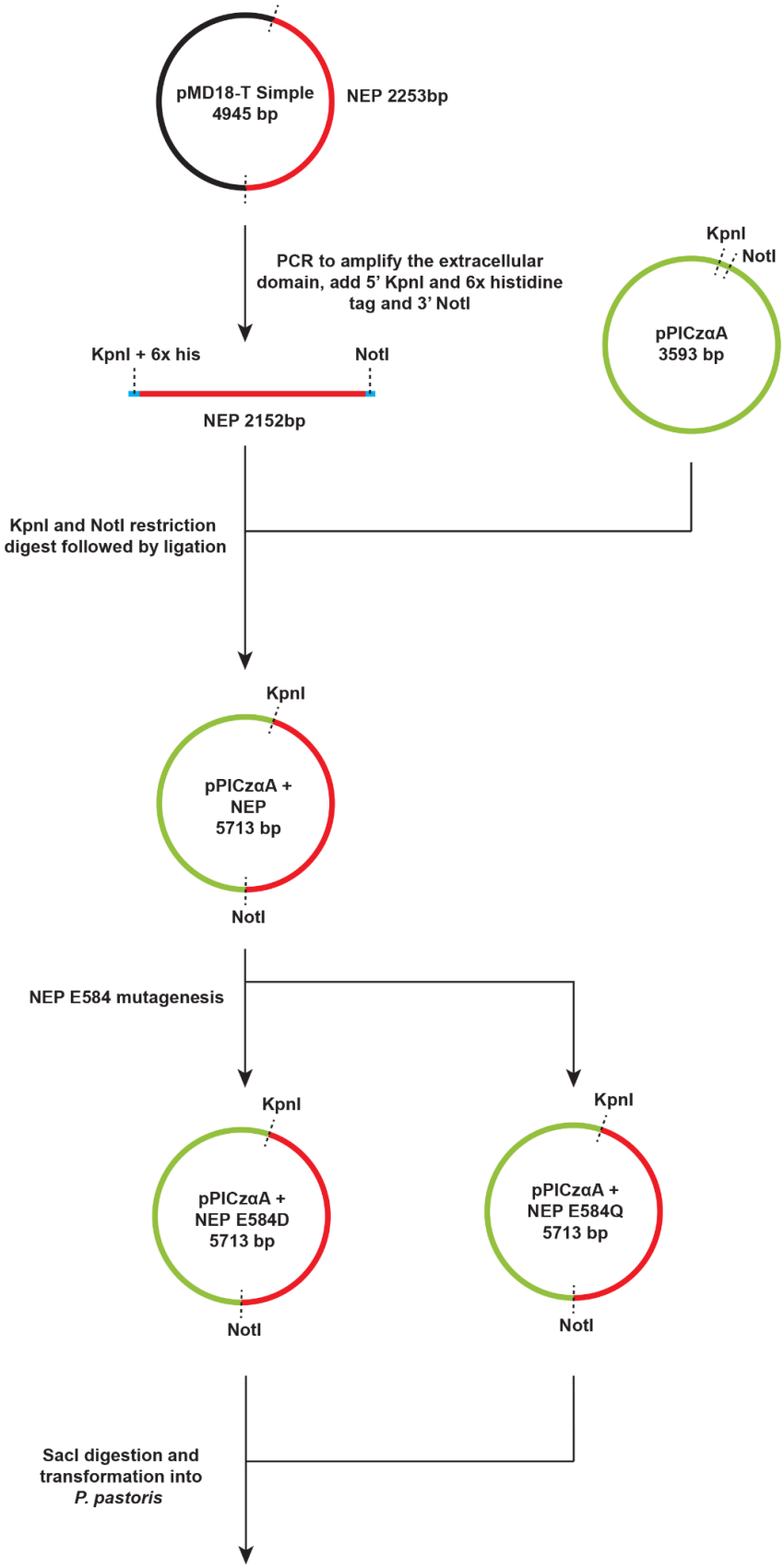


Figure 5.12, Cloning plan for generation of active site mutant NEP. pPICzαA vectors are shown with the extracellular domain of NEP with point mutations E584D or E584Q (NEP E584D pPICzαA and NEP E584Q pPICzαA).

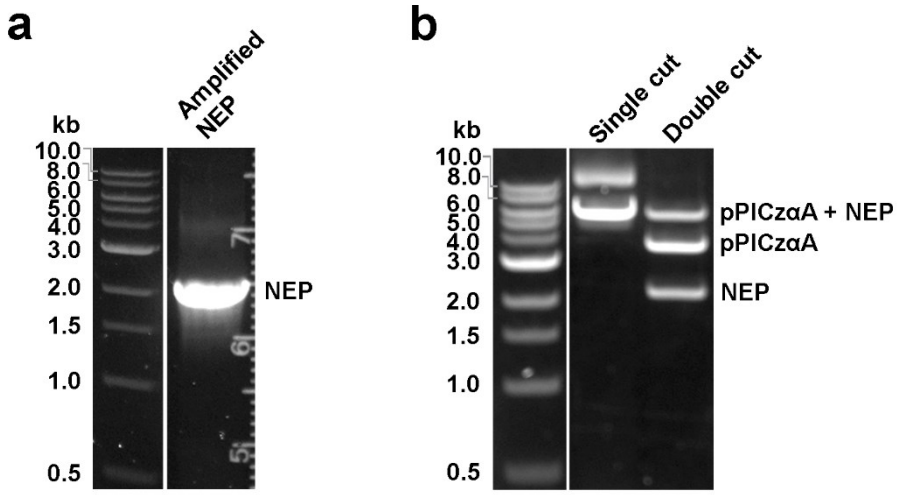


Figure 5.13, Generation of NEP pPICzαA. (a) DNA product from a PCR reaction using full length NEP pMD18-T simple as the DNA template and primers Amplification 1 Fwd and Rev. The resulting product was 2152bp and contained the extracellular domain of NEP with a 5' KpnI site, 6x histidine tag and a 3' NotI site. (b) Ligation product of NEP and pPICzαA cut with NotI or NotI and KpnI (digestion incomplete).

5.4.5 Mutant NEP expression in P. pastoris

P. pastoris clones transformed with NEP E584D pPICzαA or NEP E584Q pPICzαA were screened for protein expression. Often colonies can grow on a YPD zeocin plate without DNA integration due to transient expression of the transformed pPICzαA vector. Because of this, multiple clones were screened to identify a clone with the integrated construct and high levels of protein expression.

Only two *P. pastoris* clones grew on a YPD plate containing zeocin after NEP E584D pPICzαA transformation. Both clones were used for small-scale protein expressions to assess if either expressed NEP E584D protein. Post-induction supernatant samples were analysed and both clones 1 and 2 contained a band ~110 kDa observed on an SDS-PAGE gel and a western blot (Figure 5.14). As discussed in Chapter 3, while the mass of NEP is 81.8 kDa the predicted weight with glycosylation is expected to be higher. This in addition to the positive western blot result confirmed the band corresponded to NEP E584D.

SDS-PAGE analysis of the supernatant sample from *P. pastoris* clone 2 revealed several protein bands were present in addition to the ~110 kDa band (Figure 5.14a). The additional

signal on the corresponding western blot (Figure 5.14b) suggested that these additional protein bands correlated to NEP E584D degradation products. As these degradation products were not present in clone 1 it is this clone that was used for large-scale protein expression.

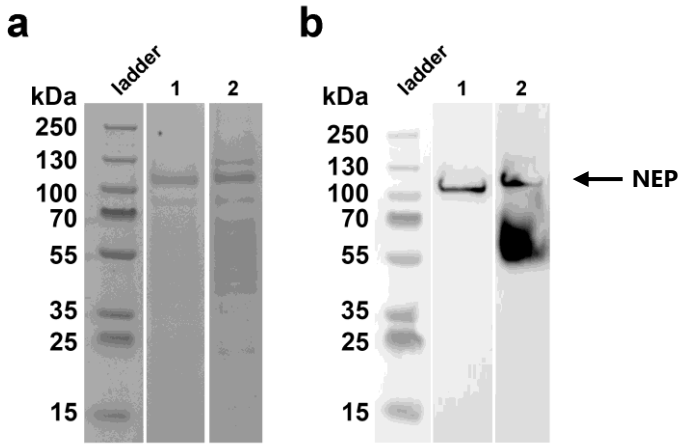


Figure 5.14, Small-scale NEP E584D expression. (a) SDS-PAGE gel electrophoresis analysis of post-induction supernatant samples from two P. pastoris clones expressing NEP E584D. (b) Western blot analysis of the same samples loaded onto the SDS-PAGE gel electrophoresis using a histidine tag antibody.

Large-scale protein expression and purification was performed using clone 1 in an identical way to wild-type NEP as described in Chapter 3. The histidine affinity purification and size exclusion purification are shown in Figure 5.15 and Figure 5.16. SDS-PAGE gel electrophoresis was used to confirm pure protein had been obtained after purification (Figure 5.17). The quantity of the purified protein was assessed and like wild-type NEP, E584D mutant yielded ~1 mg of protein for a 1 L culture.

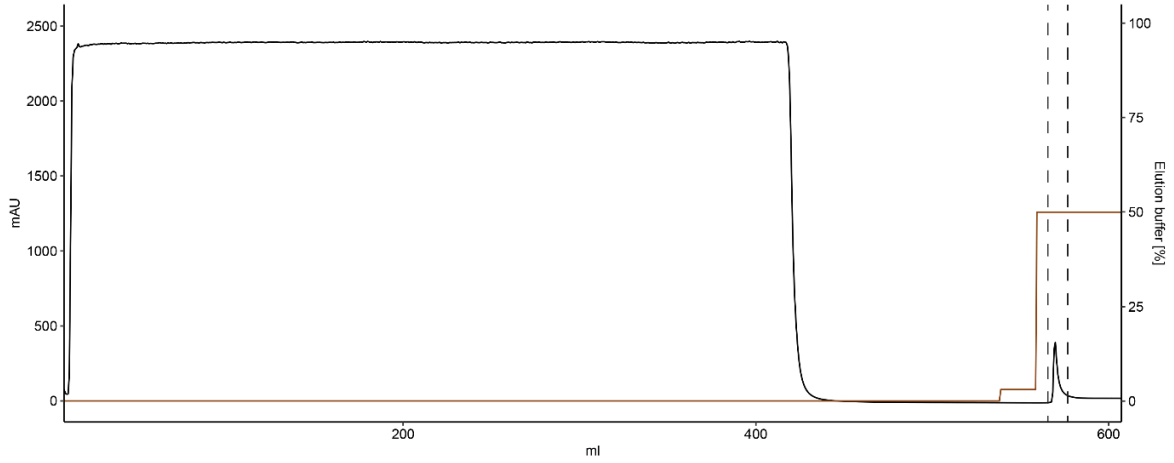


Figure 5.15, NEP E584D histidine affinity trap purification. 500 mL of protein containing supernatant from a P. pastoris culture was loaded onto a histidine affinity trap column (GE healthcare). Once loaded the column was washed with binding buffer followed by a wash step of 2% elution buffer. Protein elution was achieved at 50% elution buffer. A single peak was observed and peak containing fractions (indicated between dash lines) were pooled.

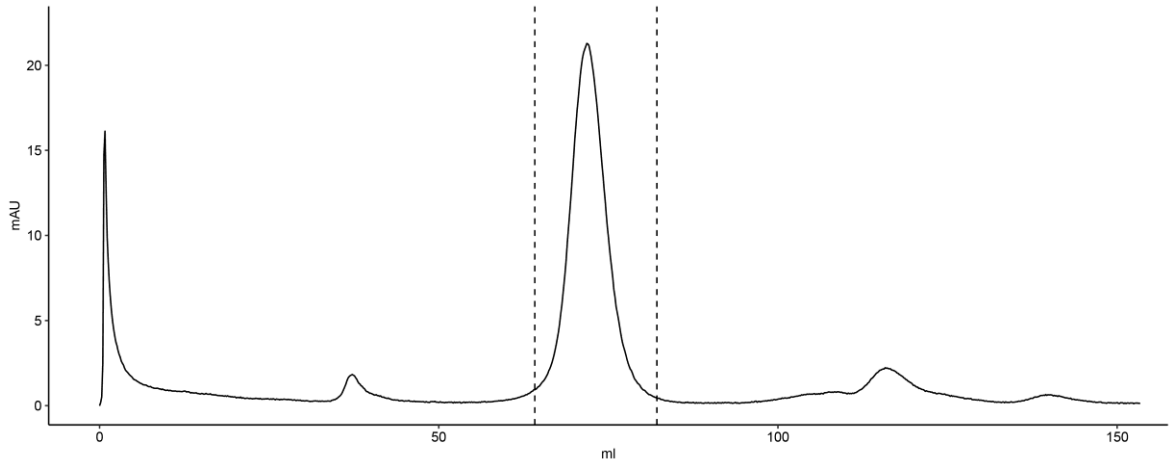


Figure 5.16, NEP E584D size exclusion purification. The pooled fractions eluted from the histidine affinity trap column were loaded onto a Superdex 16/60 size exclusion column to remove impurities from the induction media and to buffer exchange the protein into binding buffer. A single peak was observed and peak containing fractions (indicated between dash lines) were pooled.

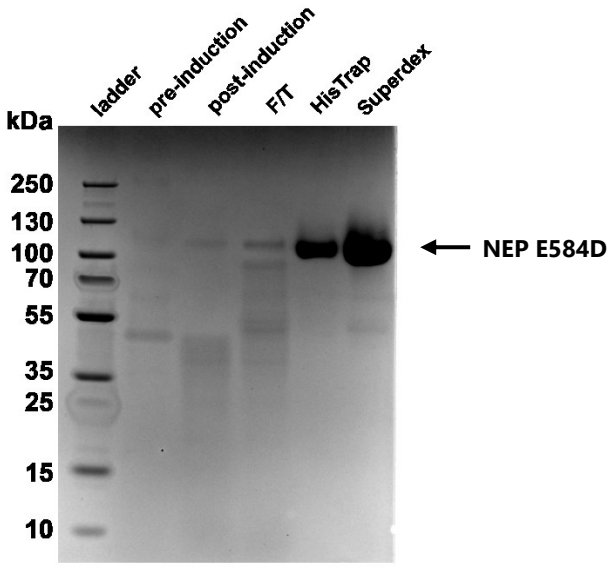


Figure 5.17, P. pastoris NEP E584D expression and purification. SDS-PAGE electrophoresis was used to analyse samples taken during NEP E584D expression. Supernatant pre- and post-induction samples were taken from the P. pastoris expression culture at 24-hours pre-induction and 72-hours post induction. Samples of the HisTrap column flow through (F/T) are also shown. The eluted and pooled fractions from the HisTrap and size exclusion purification steps reveal the high purity of the final product.

Far more colonies were present on a YPD plate containing zeocin, after NEP E584Q pPICzαA transformation. 11 of these clones were screened for protein expression on a small-scale. Post-induction supernatant samples from the 11 clones were analysed with SDS-PAGE gel electrophoresis and western blot. The SDS-PAGE analysis revealed all clones contained several protein bands (Figure 5.18a), however only clones 8 and 10 had an observable band on the western blot at ~110 kDa corresponding to NEP E584Q (Figure 5.18b). Both of these colonies were tested for large-scale protein expression.

Large-scale expression of clones 8 and 10 were performed and the expressed protein was purified using the same method for wild-type NEP and NEP E584D. SDS-PAGE electrophoresis was used to confirm pure protein had been obtained after purification (Figure 5.19a and b). While pure protein had been produced from clones 8 and 10, the quantity was much lower compared to expression of wild-type NEP and NEP E584D mutant (<0.5 mg). For this reason, kinetic and crystallisation work was performed using NEP E584D mutant protein.

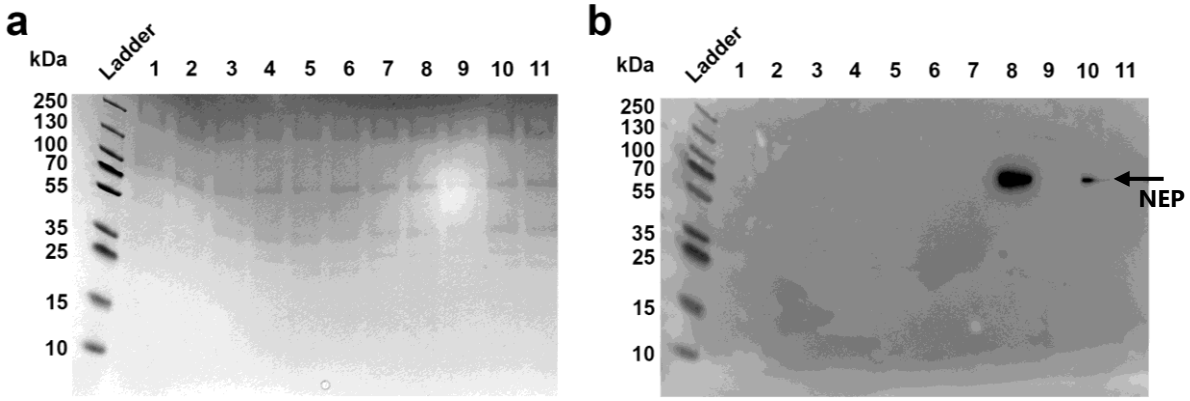


Figure 5.18, Small-scale NEP E584Q expression. (a) SDS-PAGE gel electrophoresis analysis of post-induction supernatant samples from 11 P. pastoris clones expressing NEP E584D. (b) Western blot analysis of the same samples loaded onto the SDS-PAGE gel electrophoresis.

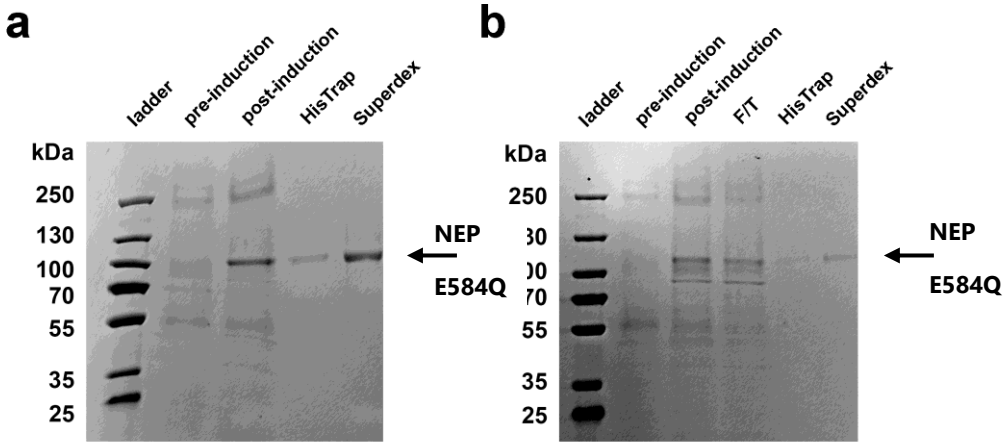


Figure 5.19, NEP E584Q expression and purification from clones 8 and 10. SDS-PAGE electrophoresis was used to analyse samples taken during NEP E584Q expressions. Supernatant pre- and post-induction samples were taken from the P. pastoris (a) clone 8 and (b) clone 10 expression cultures at 24-hours pre-induction and 72-hours post induction. The eluted and pooled fractions from the HisTrap and size exclusion purification steps reveal the high purity of the final product in both clones.

5.4.6 Protease activity of wild-type and E584D mutant NEP

Protease activity of the purified NEP E584D mutant protein was assessed using the cleavage assay as described in Chapter 3. Initial enzymatic rate was calculated for five substrate concentrations. A non-linear regression was then used to describe the experimental data (Figure 5.20). RFU were converted to molar using the '1 μM = 2103 RFU' relationship

calculated from a standard curve of product fluorescence versus product concentration (Figure 5.21). Estimates the calculated kinetic parameters K_m and V_{max} for wild-type and mutant E584D NEP are shown in Table 5.10.

The K_m and V_{max} for wild-type NEP were 38.14 µM and 0.0118 nmol/s respectively. The kinetic data for NEP E584D mutant revealed a complete absence of protease activity. It should be noted that low level activity undetectable over the assay durations may remain but based on the kinetic data collected NEP E584D was considered appropriate for substrate cocrystallisation trials.

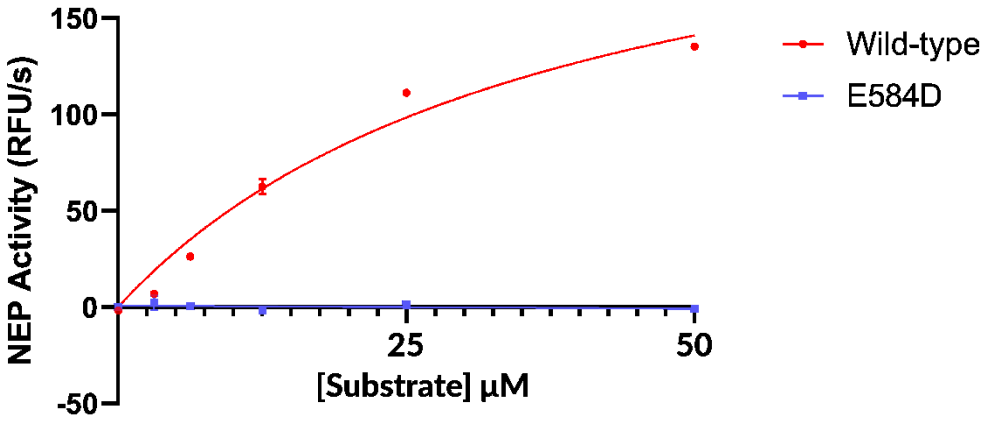


Figure 5.20, Protease activity of wild-type and mutant E584D NEP. Initial rates of activity using 0.256 mg of protein in a 100 µL reaction volume (3.125 nM) were assessed at substrate Mca-RPPGFSAFK-(Dnp) concentrations of 50, 25, 12.5, 6.25 and 3.125 µM. Initial rate values were calculated from three repeats and the mean values is plotted with ±SD error bars.

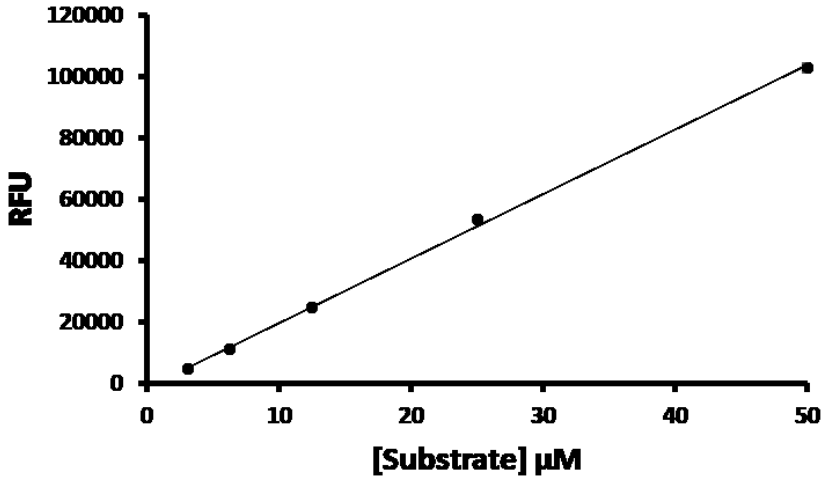


Figure 5.21, Standard curve of product fluorescence versus product concentration for Mca-RPPGFSAFK-(Dnp). Substrate concentrations were 50, 25, 12.5, 6.25 and 3.125 µM and maximal fluorescence at each concentration was recorded. A line of best fit was used to determine the relationship between substrate concentration and RFU (1 µM = 2103 RFU). For the 100 µL reaction volume a ratio of 0.1 nmoles = 2103 RFU was calculated. Maximal fluorescence at each substrate concentration was repeated three times with the mean and ±SD error bars shown above.

	Wild-type NEP	**NEP E584D mutant**
V_{max} **(nmol/s)**	0.0118 (0.0097 – 0.0153)	-
K_m **(µM)**	38.14 (26.00 – 59.45)	-

Table 5.10, Kinetic parameters for wild-type NEP and NEP E584D. Approximations for V_{max} and K_m were calculated from a non-linear regression in Graph Pad Prism using the data from Figure 5.20. 95% confidence limits are shown in parentheses. Absence of V_{max} and K_m values for NEP E584D are due to no calculatable activity. Activity was determined using 0.256 mg of NEP or NEP E584D in a 100 µL reaction volume (3.125 nM).

5.4.7 Crystallisation and optimisation of NEP E584D

Crystallisation of NEP E584D mutant was initially performed using the same crystallisation condition used for wild-type NEP (JCSG-plus A12, 0.2 M KNO3, 20% (v/v) PEG 3350). Slight variation in this condition was implemented with the aim of increasing speed of

crystallisation or generating crystal that consistently diffracted to higher resolution. Variation was done by screening 80% (v/v) of condition JCSG A12 with 20% (v/v) JCSG-plus or PACT premier. Many of the new crystallisation conditions generated crystals. Cocrystallisation and soaking experiments were performed with new conditions using NEP E584D with substrates including C-type natriuretic peptide, B-type natriuretic peptide, bradykinin, fluorescent Mca-RPPGFSAFK-(Dnp) peptide and amyloid-β fragments 1-40, 10-16, 1-16 and 35-42. The following published manuscript reports the highest-resolution structure with the best peptide-bound density for NEP E584D mutant obtained through crystal screening with the listed substrates.

5.5 Published article

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS
Apr 13, 2020

This Agreement between University of Bath -- Stephen Moss ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4796040893122
License date	Mar 25, 2020
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	FEBS Letters
Licensed Content Title	Crystal structure of peptide-bound neprilysin reveals key binding interactions
Licensed Content Author	Stephen Moss, Vasanta Subramanian, K. Ravi Acharya
Licensed Content Date	Sep 21, 2019
Licensed Content Volume	594
Licensed Content Issue	2
Licensed Content Pages	10
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title	Structure and functional studies of key amloid degrading enzymes
Expected presentation date	Apr 2020
Requestor Location	University of Bath

This declaration concerns the article entitled:									
Crystal structure of peptide-bound neprilysin reveals key binding interactions									
Publication status (tick one)									
draft manuscript		**Submitted**		**In review**		**Accepted**		**Published**	✓
Publication details (reference)	**Moss S,** Subramanian V & Acharya KR (2019) Crystal structure of peptide-bound neprilysin reveals key binding interactions								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate contributed to/ considerably contributed to/predominantly executed the… Research design: KRA Methodology design: SM Protein purification and expression of NEP: SM Assessment of activity: SM Crystallisation: SM Structural determination: SM Analysis of data: SM, KRA Wrote the manuscript: SM Prepared the figures: SM Edited the manuscript: SM, VS, KRA								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed	**Stephen Moss**					**Date**	14/04/20		

FEBS Letters

FEBS PRESS science publishing by scientists

Crystal structure of peptide-bound neprilysin reveals key binding interactions

Stephen Moss, Vasanta Subramanian and K. Ravi Acharya

Department of Biology and Biochemistry, Claverton Down, University of Bath, UK

Correspondence
K. Ravi Acharya, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK
Tel: +44 1225 386238
E-mail: bsskra@bath.ac.uk

(Received 6 August 2019, revised 4 September 2019, accepted 5 September 2019, available online 21 September 2019)

doi:10.1002/1873-3468.13602

Edited by Miguel De la Rosa

Neprilysin (NEP) is a promiscuous zinc metalloprotease with broad substrate specificity and cleaves a remarkable diversity of substrates through endopeptidase action. Two of these – amyloid-β and natriuretic peptides – implicate the enzyme in both Alzheimer's disease and cardiovascular disease, respectively. Here, we report the creation of a catalytically inactive NEP (E584D) to determine the first peptide-bound crystal structure at 2.6 Å resolution. The structure reveals key interactions involved in substrate binding which we have identified to be conserved in other known zinc metalloproteases. In addition, the structure provides evidence for a potential exosite within the central cavity that may play a critical role in substrate positioning. Together, these results contribute to our understanding of the molecular function of NEP.

Keywords: crystallography; NEP; neprilysin; neutral endopeptidase; peptide-bound; protein structure; zinc metalloprotease

Neprilysin (NEP, EC 3.4.24.11) is a transmembrane M13 zinc metalloprotease that exhibits endopeptidase activity on over 30 short peptide substrates [1,2]. Recently, NEP has received considerable attention for its role in amyloid-β degradation and represents a potential therapeutic target in Alzheimer's disease research [3]. The enzyme is, however, better characterised for its role in blood pressure regulation through the degradation of several small peptide vasodilators which include bradykinin, substance P and natriuretic peptides [4].

The mechanism of degradation of three main classes of natriuretic peptides – atrial, B-type and C-type (ANP, BNP and CNP) by NEP are well documented [5]. These three classes of peptides have been studied for their promotion of vasodilation mediated through interaction with membrane guanylate cyclases and natriuretic receptors [6]. Whilst each peptide varies in length (between 22 and 32 amino acids), all share a common 17-residue central ring structure with 10 of the 17 residues conserved [6]. Differences in amino acid sequence cause varied susceptibility to NEP degradation as follows: CNP>ANP>BNP [7]. Their degradation and inactivation by NEP result in vasoconstriction and elevated blood pressure levels. As such, NEP has been targeted for inhibition in order to elicit an antihypertensive response in patients with high blood pressure [8].

The molecular structure of NEP is comprised of a 27-residue intracellular domain, followed by a 23-residue transmembrane domain and finally a 699-residue extracellular catalytic domain. To date, several crystal structures of the NEP extracellular domain have been reported for both substrate-free [9] and inhibitor-bound forms [10–15]. These structures reveal the domain is ellipsoid in shape with the active site contained within a large occluded cavity. The active site is based around the conserved metalloprotease HEXXH motif where two conserved histidine residues (His583 and His587) coordinate a zinc ion and the

Abbreviations
CNP, C-type natriuretic peptide; ECE-1, Endothelin-converting enzyme I; IDE, Insulin-degrading enzyme; NEP, Neprilysin; PEG, polyethylene glycol; PreP, presequence peptidase.

conserved glutamate (Glu584) is responsible for polarisation of a solvent molecule which then performs nucleophilic attack on the substrate's scissile peptide bond [16].

It should be noted that NEP and a few other zinc metalloproteases – including endothelin-converting enzyme I (ECE-1) [17], insulin-degrading enzyme (IDE) [18] and presequence peptidase (PreP) [19] – are classed as cryptidases due to the location of the active site within a large cavity [20]. Each cryptidase must undergo a conformational change in which the structure transitions from open, where substrates can access the cavity, to closed, where substrates are trapped, so that substrate recognition can occur [20]. Cryptidases range in size from 85 to 120 kDa and can degrade many peptide substrates that vary in sequence, size, shape and surface charge.

The promiscuous nature of NEP, and other cryptidases, raises questions as to their mechanism of substrate specificity. NEP inhibitor-bound crystal structures have provided some insights into the substrate specificity, enabling the characterisation of S1, S1' and S2' [21] binding pockets at the catalytic site. S1' has been reported to be the largest contributor to substrate specificity having preference for hydrophobic amino acids. Specificity is also achieved through the limited cavity volume imposing constraints on maximal substrate size. For NEP, the cavity volume of approximately 5000 $Å^3$ results in the selection of peptides under 5 kDa in size [22].

Despite the availability of a number of NEP crystal structures, a peptide-bound complex has yet to be reported. This can be mainly attributed to rapid cleavage of peptides by NEP, hence not ideal for X-ray crystallography studies investigating protein–substrate binding interactions. Moreover, crystal growth, and subsequent data collection, of protein crystals requires lengthy periods of incubation, and NEP would process any peptide substrate long before data collection could take place. In this study, a mutant NEP E584D protein was expressed and confirmed to be catalytically inactive. This protein was used in the crystallographic study to prevent peptide cleavage and allow binding interactions to be preserved.

Here, we report the structures of catalytically inactive NEP E584D in a substrate-free and peptide-bound form. In the peptide-bound form, a C-type natriuretic peptide (CNP) is bound to the active site, revealing main chain binding interactions and providing insights into the general mechanism of peptide interaction. These results contribute to the understanding of a key enzyme NEP that has a role to play in both Alzheimer's disease and cardiovascular disease.

Materials and methods

All reagents used were purchased from Sigma-Aldrich Company Ltd (Gillingham, UK) or Fisher Scientific UK Ltd (Loughborough, UK) unless otherwise specified.

Plasmid construction and transfection

cDNA encoding full-length human NEP in cloning vector pMD18-T Simple was purchased from Sinobiology. PCR was used to amplify the extracellular domain of NEP (Thr49-Trp749), add a 5' KpnI site and 6× histidine tag and add a 3' NotI site. The amplified PCR product and vector pPICzαA were digested with KpnI and NotI (NEB). The PCR product was then ligated into pPICzαA downstream of a secretion tag. This tag was transcribed with NEP for secreted protein expression but was cleaved after secretion.

Mutation of the NEP residue 584 from glutamate to aspartate (E584D) was conducted using PCR site-directed mutagenesis (NEB Q5) and confirmed using DNA sequencing. The NEP pPICzαA construct was integrated into the genome of *Pichia pastoris* GS115 in accordance with the Invitrogen EasySelect manual (cat. no. K1740-01) and using restriction enzyme SacI (NEB). A full list of primer sequences is given in Table 1.

Protein expression and purification

Expression of recombinant NEP and NEP E584D extracellular domains from the *P. pastoris* GS115 was performed in accordance with the Invitrogen EasySelect manual. Briefly, a preinduction culture was incubated for 24 h in Buffered Glycerol-complex Medium to generate the desired cell density. Media was changed to Buffered Methanol-complex Medium for a 72-h induction, and 100% methanol was added at 24 and 48 h to maintain the methanol concentration.

After induction, the supernatant was harvested by centrifugation at 6000 RCF for 20 min at 4 °C. Trizma and NaCl were added to the supernatant to give final concentrations of 25 mM Trizma and 150 mM NaCl. The supernatant was then centrifuged at 6000 RCF for 20 min at 4 °C and filtered through a 0.22 μm filter (Millipore UK Limited, Watford, UK).

Table 1. Primer sequences used for NEP E584D clone creation.

	Primer sequence
Cloning in pPICzαA	
Forward	CGCTGGTACCCATCACCACCACCATCACGAAT TCCGGACCTACGATGATGGTATTTGCAAG
Reverse	GCTTGCGGCCGCTCACCAAACCCGGCACTTC
Mutagenesis	
Forward	CATAGGACACCAAATCACCCATGGCTTCG
Reverse	ACCATGCCGATGCCCCCA

The supernatant was loaded onto a 5 mL HisTrap affinity column (GE Healthcare Bio-Sciences, Pittsburgh, PA, USA) at 5 mL·min^{-1} pre-equilibrated with Buffer A (25 mM Trizma, 150 mM NaCl and 2 mM $MgCl_2$, pH 7.5). After loading, the column was washed with 2% Buffer B (25 mM Trizma, 150 mM NaCl, 2 mM $MgCl_2$ and 500 mM imidazole, pH 7.5) until the UV trace returned to baseline. NEP was eluted using a single 50% Buffer B wash. Protein-containing fractions were pooled and concentrated to 1 mL using a 30 kDa Millipore centrifuge filter. The concentrated NEP was loaded onto a 16/60 Superdex HiLoad 200 column for size exclusion. Again, protein-containing fractions were pooled and stored at 4 °C for short-term storage or −20 °C for longer-term storage. SDS/PAGE was used to confirm the presence and purity of NEP after purification.

Enzyme assay

The activities of NEP and NEP E584D were assessed by measuring cleavage of the fluorogenic peptide Mca-RPPGFSAFK-(Dnp) [Enzo Life Sciences (UK), Exeter, UK] [23]. Protein was diluted to a concentration of 31.3 nM in Buffer A, and 50 µL of diluted protein was added to a black 96-well microplate (cat. no. 655076; Greiner Bio-One Ltd., Stonehouse, UK). Fifty microlitres of 100 µM Mca-RPPGFSAFK-(Dnp), dissolved in Buffer A with 10% DMSO, was added to a black 96-well microplate (Greiner Bio-One) to initiate the reaction.

Each fluorescence assay was conducted in triplicate. Fluorescence was recorded at 20-s intervals for a length of 35 min on a CLARIOstar high-performance microplate reader (BMG LABTECK Ltd., Aylesbury, UK). Excitation was achieved at 328 nm and emission was detected at 393 nm [9].

Circular dichroism

The secondary structure composition of NEP and NEP E584D was analysed using CD. Protein samples of NEP and NEP E584D were diluted to 2.725 µM and 2.6 µM, respectively, before analysis. Data were collected in triplicate using a 1 mm quartz cuvette and a Chirascan CD spectrometer at 20 °C. The wavelength scanning range was between 200 and 300 nm with a 1 nm step size, monochromator bandwidth of 3 nm and sample time-per-point of 3 s.

Crystallisation

Protein crystallisation was performed using sitting-drop vapour-diffusion technique. NEP E584D (8 mg·mL^{-1}) was dispensed with a Phoenix crystallisation robot (Art Robbins Instruments, Sunnyvale, CA, USA) into an Intelli-Plate 96 (cat. no. 102-0001-03; Art Robbins Instruments) with protein to crystallisation buffer ratio of 0.2 : 0.2. Crystallisation buffer for the substrate-free structure was 0.16 M potassium nitrate, 0.04 M sodium bromide, 0.02 M bis-tris propane and 20% (w/v) polyethylene glycol 3350. Crystallisation plates were incubated at 18 °C and crystals were formed within 2 months.

Attempts to co-crystallize NEP E584D with peptide were unsuccessful. Instead, the peptide-bound structure was achieved through soaking. Crystals for soaking were generated using crystallisation buffer composed of 0.16 M potassium nitrate, 0.04 M sodium iodide and 20% (w/v) polyethylene glycol 3350. Again, crystallisation plates were incubated at 18 °C and crystals were formed within 2 months. Once the crystals were formed, 0.2 µL of 9.1 mM CNP was added to the crystallisation drops and incubated for further 14 days at 18 °C before data collection. The CNP used had the amino acid sequence 'GLSKGCFGLKLDRIGSMSGLGC' and was sourced from GenScript (Cas. No. 127869-51-6). All crystals used for data collection were flash-cooled in liquid nitrogen for storage without the use of cryoprotectant.

X-ray data collection and refinement

X-ray diffraction data collection for the NEP E584D substrate-free and peptide-bound structures was conducted on Diamond Light Source (Didcot, Oxford, UK) beamlines I04 (mx17212-37) and I04 (mx17212-46), respectively. At the time of collection, both beam lines were fitted with Eiger2 X 16M detectors.

X-ray diffraction data were indexed and integrated using DIALS [24]. Data were scaled and merged using AIMLESS [25,26] where resolution cut-offs were applied. The structures were solved by molecular replacement using Phaser [27] and substrate-free NEP (6GID) [9]. Programs Refmac5 [28] and Coot [29] were used for refinement of the structures.

Validation of the structures was achieved using programs MolProbity [30] and PDB validation [31]. All structural images and superpositions were generated in CCP4mg [32]. The final refined structures were deposited in the PDB under codes 6SH1 and 6SH2.

Results and Discussion

Catalytic inactivation of NEP

To investigate peptide-binding interactions, it was necessary to use catalytically inactive NEP. Inactive NEP could not perform peptide cleavage and hence allowed protein–peptide interactions to be preserved in crystallisation experiments. Similar crystallographic approaches have successfully been reported in the study of protein–peptide binding interactions with other catalytically inactive zinc metalloproteases [18,33].

Catalytic inactivation of NEP was achieved by mutation of active site glutamate 584 to an aspartate (E584D). DNA sequencing, as described in materials and methods, was used to confirm the presence of this mutation. The extracellular domain of NEP E584D was then expressed in *P. pastoris*. In the following results section, the extracellular domain of NEP E584D will be referred to as just NEP E584D. Activity of NEP E584D was measured using a fluorescence assay (Fig. 1). Unlike wild-type NEP, NEP E584D showed loss of catalytic activity. Whilst no activity for NEP E584D was observed over the 35-min assay, there remained the possibility that low-level residual activity may still be present.

CD was used to compare the secondary structure composition of NEP and NEP E584D. The high degree of similarity in CD traces confirmed that mutation E584D did not induce secondary structure variation (Fig. 2). As such, NEP E584D protein was considered suitable for crystallisation experiments.

Comparisons of NEP E584D and wild-type NEP substrate-free crystal structures

X-ray crystallography data for substrate-free NEP E584D were collected, and the crystal structure (6SH1) was determined in the monoclinic space group $P2_1$ to a resolution of 2.1 Å. It should be noted that several lower-resolution X-ray crystallography data sets for substrate-free NEP E584D were collected and solved in the more frequently observed $P3_221$ space group. Details of data collection and refinement statistics are listed in Table 2. The substrate-free structure of NEP E584D was used to determine whether the presence of mutation E584D caused significant structural perturbation from wild-type NEP.

Previously, we had reported the structure of wild-type substrate-free NEP (6GID) [9] to a resolution of 1.9 Å. This wild-type structure adopted a closed conformation and revealed the position of active site residues in a substrate-free state. Wild-type substrate-free NEP (6GID) was superimposed onto substrate-free NEP E584D (Fig. 3A). A RMSD value of 0.57 Å, calculated using a least squares comparison for α-carbon residues 54–749, confirmed the overall structure had been preserved. The high degree of structural similarity also confirmed that variation in space group between $P3_221$ and $P2_1$ did not cause structural changes due to crystal packing. NEP E584D remained in a closed conformation and contained the same four glycosylation sites. Electron density for the zinc ion remained clear, and only minimal movement was present in residues surrounding or involved in zinc coordination (Fig. 3B). The only notable difference between wild-type and E584D substrate-free NEP structure was the molecule that provided additional coordination to the zinc ion. In wild-type substrate-free NEP, a phosphate ion coordinates the zinc ion whilst in substrate-free NEP E584D a single water molecule coordinates the zinc ion. The similarities between wild-type and mutant structures supported the view that NEP

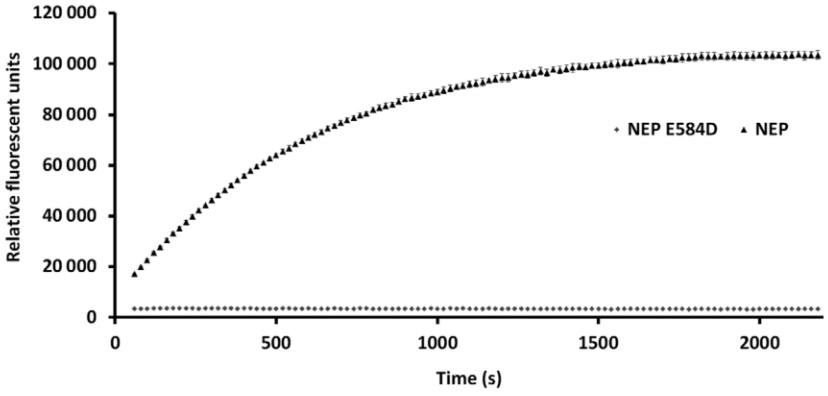


Fig. 1. Enzymatic assay confirming NEP E584D inactivation. NEP or NEP E584D was incubated with reporter peptide Mca-RPPGFSAFK(Dnp). Cleavage of the reporter peptide produces detectable fluorescence which is recorded over time. Above, assays for Mca-RPPGFSAFK(Dnp) cleavage by NEP or NEP E584D are shown. The average of three assays has been plotted at each data point with standard error bars.

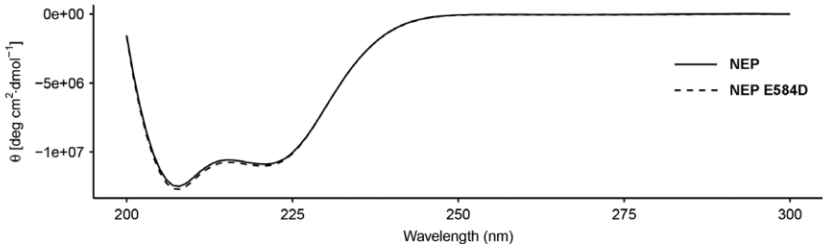


Fig. 2. CD comparison between wild-type NEP and NEP E584D. The overlapping traces confirm NEP E584D has the same secondary structure composition as wild-type NEP and are indicative of a largely α-helical protein.

Table 2. NEP E584D X-ray crystallography and refinement statistics. Values in parentheses are for highest-resolution shell.

	Substrate-free	Peptide-bound
Beamline	I04, DLS	I03, DLS
Wavelength (Å)	0.9795	0.9762
Crystallographic statistics		
Space group	$P2_1$	$P3_221$
Unit cell dimensions		
a,b,c (Å)	76.14, 99.75, 100.12	109.27, 109.27, 112.58
α, β, γ (°)	90.00, 106.23, 90.00	90.00, 90.00, 120.00
Resolution range (Å)	69.22–2.10 (2.14–2.10)	72.44–2.60 (2.72–2.60)
R_{merge}	0.178 (3.676)	0.144 (6.977)
$CC_{1/2}$	0.992 (0.464)	1.000 (0.604)
$\langle I/\sigma\ (I)\rangle$	6.6 (1.0)	17.3 (1.0)
Completeness (%)	99.9 (99.8)	100 (100)
No. observed reflections	576297 (32234)	973293 (107982)
No. unique reflections	83834 (4581)	24386 (2945)
Multiplicity	6.9 (7.0)	39.9 (36.7)
Refinement statistics		
R_{work}/R_{free} (%)	21.94/26.35	24.43/28.82
Rmsd		
Bond lengths (Å)	0.003	0.002
Bond angles (°)	1.212	1.148
Ramachandran statistics		
Favoured (%)	97	96
Allowed (%)	3	4
Outliers (%)	0	0
Wilson B-factor ($Å^2$)	39.9	84.3
Average B-factors ($Å^2$)		
Protein atoms	44.54	102.18
Ligand atoms	–	104.16
Water molecules	42.67	62.68
Zinc ion	34.10	81.62
No. of atoms		
Protein	11226	5512
Ligand	–	19
Water	495	1
Zinc ion	2	1
PDB code	6SH1	6SH2

E584D was an appropriate model to study peptide-binding interactions. Fourier electron density maps of residue 584 in NEP and NEP E584D highlight the presence of the mutation (Fig. 3C).

Peptide-bound crystal structure of NEP E584D

The peptide-bound crystal structure of NEP E584D (6SH2) was determined in the trigonal space group $P3_221$ to a resolution of 2.6 Å. The active site of NEP E584D, contained within the large central cavity, had sufficient electron density to model a four-residue peptide product from the 22-residue long CNP (Fig. 4). These modelled residues occupied the S1, S1', S2' and S3' subsites that were previously characterised in inhibitor-bound structures [10–15].

Previous high-performance liquid chromatography data for CNP has reported that at least 6 cleavage sites are targeted by NEP [34]. The ability of NEP to cleave CNP at several distinct positions dictates that in a single protein crystal, the interaction of the residues of CNP with the active site will be different in individual NEP molecules. Therefore, from the experimental X-ray crystallography data, whilst we were able to ascertain the mode of peptide binding for the α-carbon main chain, we could not model the specific side chains that occupy the S1–S3' subsites. Perhaps with higher-resolution data, it may be possible to identify the most common residues occupying these subsites in the conditions tested.

For the best interpretation of the electron density for CNP and to provide information on the mode of peptide binding, a poly-alanine tetra-peptide was modelled bound in the NEP E584D active site (Fig. 5A). A F_o-F_c Fourier difference map showed some continuation of electron density in areas of unmodelled side chains. Additional density was observed at the N- and C-terminus of the modelled peptide chain indicating the presence of additional residues that could not be resolved. The majority of CNP could not be modelled, suggesting much of the peptide is flexible (disordered) and does not exhibit tight binding interactions with the cavity lining. This also indicates that the modelled P1, P1', P2' and P3' residues are likely to be the primary sites of interaction between NEP E584D and the peptide. An alternative possibility is that peptide cleavage could have occurred due to low levels of protease activity which went undetected whilst performing enzyme assay measurements. However, the observed electron density between P1 and P1' was continuous, indicating that no cleavage had occurred in the peptide fragment that was modelled.

Despite the inability to model the side chains of CNP, several binding interactions could be identified between the modelled poly-alanine main chain of CNP and NEP E584D (Fig. 5B). Based on our interpretation of the electron density, we suggest that Asn542 and Ala543 interact with the amino group of P1'. Asn542 also interacts with the carboxyl group of P2'. Arg717 interacts with the carboxyl group of P1', and additional interactions are provided through coordination of the zinc ion by the carboxyl group of the P1 residue. Interestingly, these main chain interactions

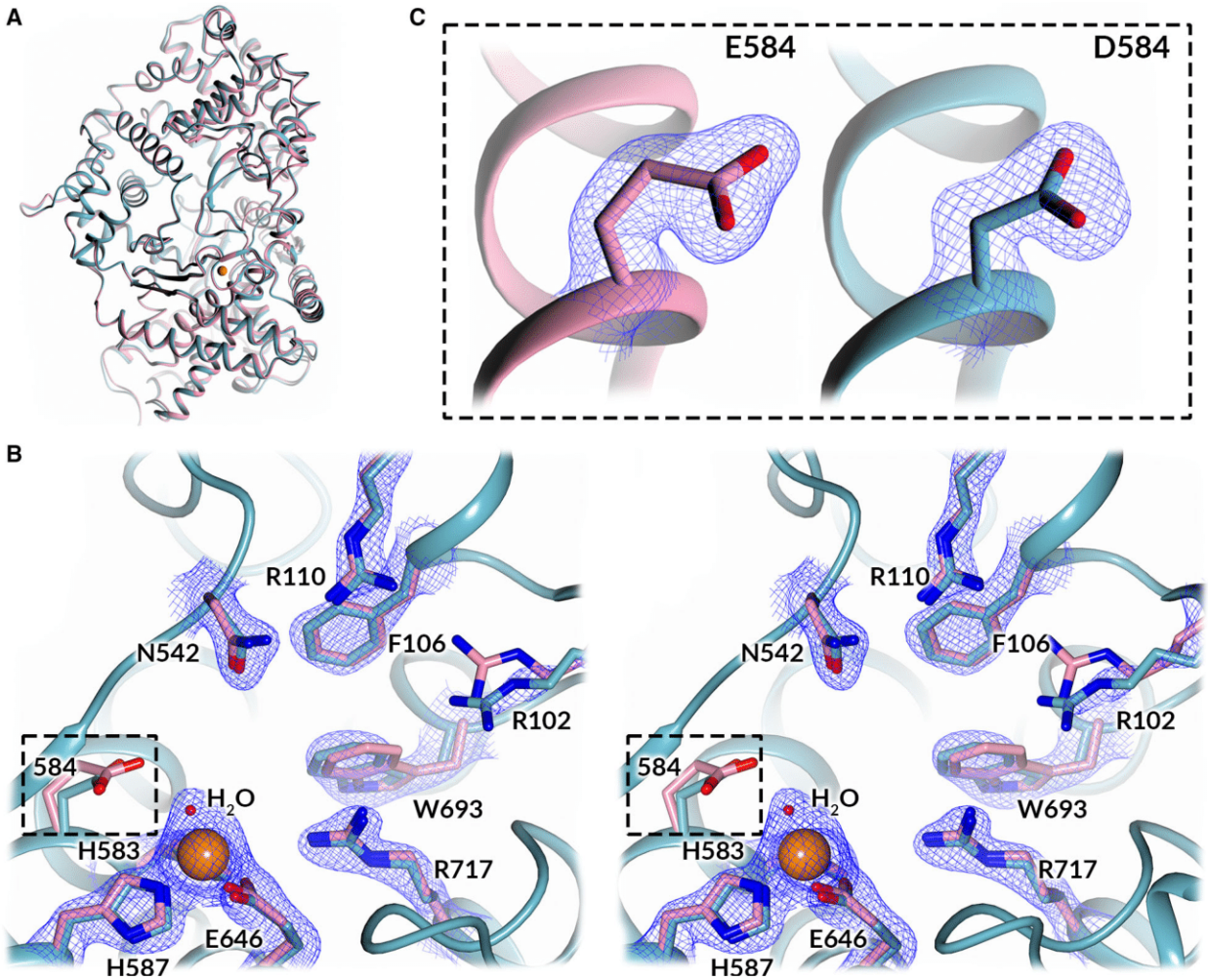


Fig. 3. Comparison of NEP and NEP E584D crystal structures. (A) A superposition of the substrate-free NEP (6GID) (pink) and NEP E584D (6SH1) (turquoise) crystal structures. (B) A stereoview of the superimposed active site residues. Electron density for residues E584 (6GID) and D584 (6SH1). (C) $2F_o$-F_c maps are set to a contour level of 1σ.

induce a > 90° bend in the CNP around the P1' residue, and their importance has been highlighted in previous inhibitor-bound structures. Phosphoramidon, a generic metallo-endopeptidase inhibitor, exhibits very similar binding interactions to the zinc ion, Asn542, Ala543 and Arg717 (Fig. 6A) [10]. This is also true of sacubitril, a NEP specific inhibitor, where all interactions except those from Ala543 are observed [15].

Further, residues Asn542, Ala543 and Arg717 are conserved in other metalloproteases including the previously mentioned cryptidases IDE, PreP and ECE-1. Peptide-bound structures for both IDE (2G47) and PreP (2FGE) reveal identical binding interactions to these conserved residues, leading to a highly similar peptide orientation. A superposition of peptide-bound NEP E584D and IDE structures (2G47) illustrates this observation (Fig. 6B). The high degree of similarity in peptide binding further validates our peptide-bound NEP E584D structure. In the absence of a peptide-bound structure of ECE-1, we predict a similar mode of peptide binding given the high structural homology to NEP.

In addition to the main chain interactions discussed above, peptide binding will also involve side-chain interactions. Whilst the side chains for CNP P1–P3' residues have not been modelled, the position of modelled β-carbons allows prediction of subsite residues that are important for protein–peptide interactions (Figs 5 and 7). The β-carbons for P1' and P2' are orientated towards NEP E584D residues; therefore, side chains are likely to have important roles in peptide binding. In contrast, P1 and P3' side chains are

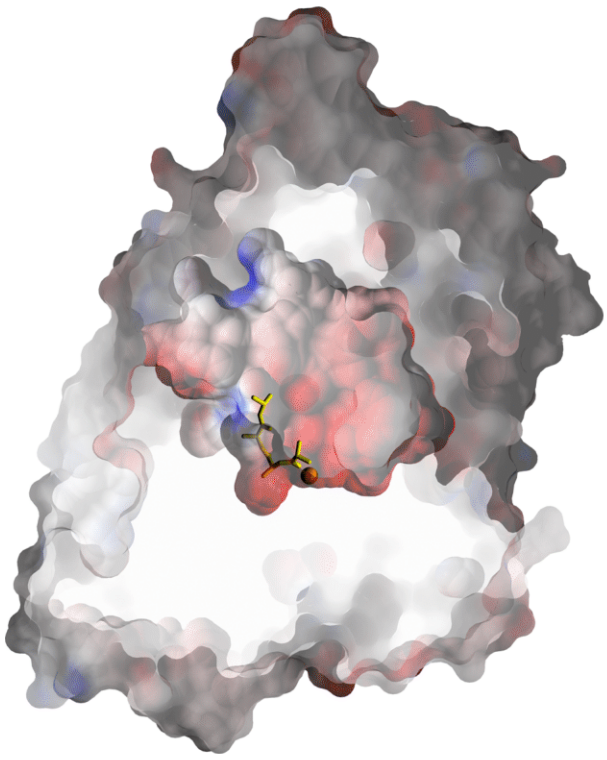


Fig. 4. Electrostatic surface representation of NEP E584D (6SH2) with bound CNP. The location of bound peptide (yellow) is shown with the catalytic zinc ion (orange).

orientated towards the cavity centre and are therefore unlikely to have close interactions with S1 and S3' sites.

The S1' and S2' sites have been characterised in inhibitor-bound structures [10–15] and exhibit several binding interactions. The S1' site is a hydrophobic pocket resulting in the reported preference of NEP to cleave substrates at the amino side of hydrophobic residues. S2' has been reported to have a preference for hydrophobic residues and can accommodate bulky side chains due to the movement involving residues Arg110, Arg102 and Phe106 [35,36].

Putative exosite

Additional unexplained electron density was observed in a proximal site approximately 14 Å away from the zinc ion in the peptide-bound NEP E584D structure (Fig. 8A). This density was completely absent in the substrate-free NEP E584D structure. It is likely that this density corresponds to a portion of CNP interacting with residues of the NEP E584D cavity. These interactions may anchor the peptide and explain why electron density is observed in this specific area. Whilst the exact region of CNP cannot be identified from electron density, the shape and size of the density can be accounted for by an amino acid residue (Fig. 8B). The modelled residue is orientated such that the electron density continues beyond the carboxyl and amino groups and the β-carbon is positioned towards polar residues Asp208, His217 and Tyr346, possibly indicating a preference for a polar residue in that location.

Similar electron density feature has been observed and reported in a peptide-bound structure of IDE [18]. IDE, like NEP, is a zinc metalloprotease that degrades several small peptide substrates and possesses a catalytic site that is located within a large central cavity. In the peptide-bound structure of IDE, electron

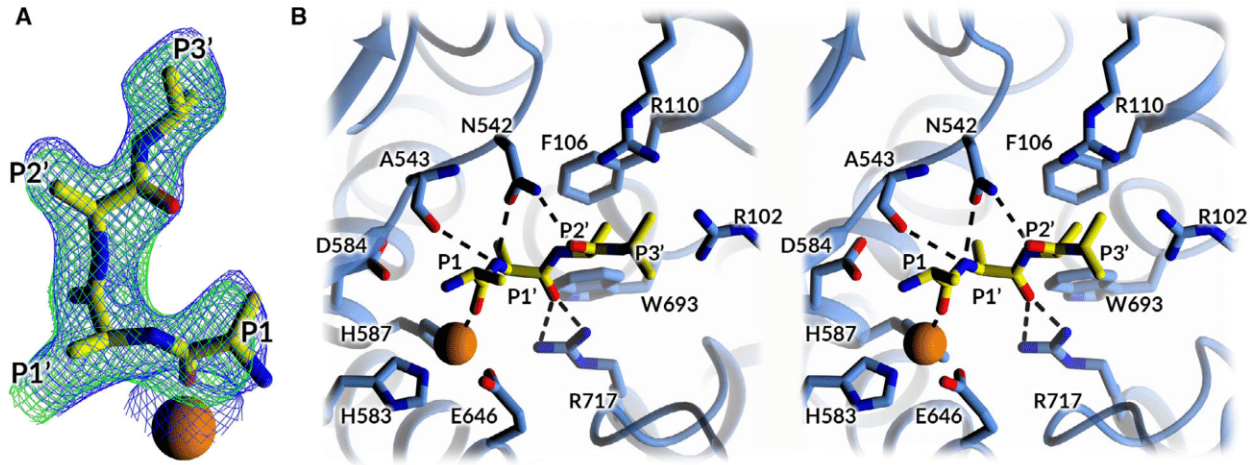


Fig. 5. CNP electron density and binding interactions. (A) The modelled peptide fragment is shown with localised $2F_o$-F_c and F_o-F_c omit electron density maps (6SH2). $2F_o$-F_c and F_o-F_c maps are set to contour levels of 1σ and 3σ, respectively. (B) A stereoview of the modelled peptide fragment with surrounding active site residues. Close range interactions are marked with dashed lines.

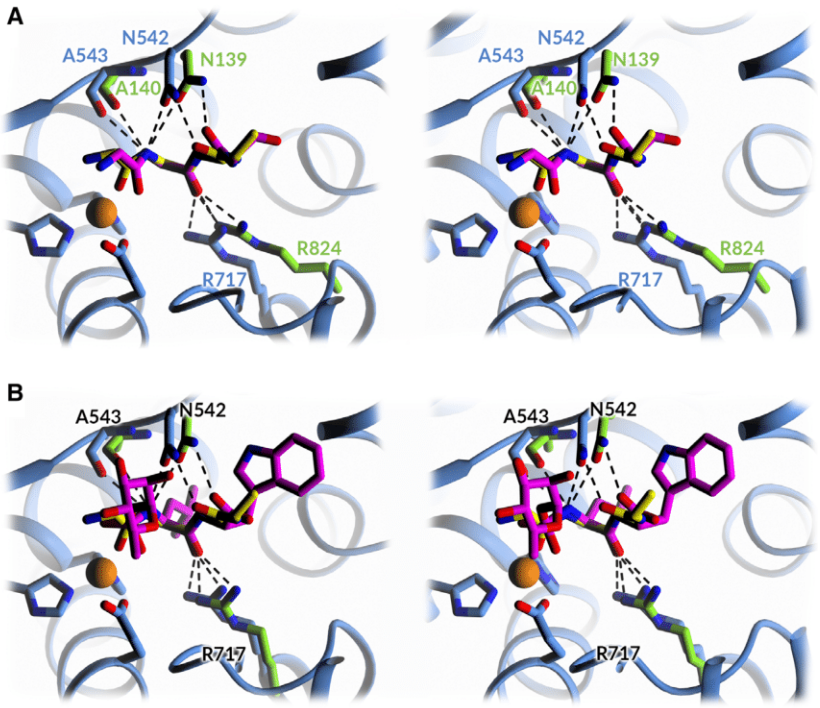


Fig. 6. Comparison of conserved active site residues and their role in substrate/inhibitor binding. Stereoviews of peptide-bound NEP E584D (6SH2) (blue) superimposed onto (A) the phosphoramidon-bound NEP structure (1DMT) (green) and (B) peptide-bound IDE structure (2G47) (green) are presented. The P1–P3′ residues of the bound peptide in NEP E584D are shown in yellow. Phosphoramidon and IDE-bound peptide are shown in magenta. Close range interactions between peptide and protein are marked with dashed lines.

density was observed for residues around the catalytic zinc in addition to the proximal site. This proximal or exosite in IDE has been proposed to be important for positioning the peptide such that the correct residues are localised to the active site.

Whilst the peptide-bound structure reported here provides a strong indication for the presence of an exosite, complete certainty in the orientation of the interacting amino acid residue is lacking. Therefore, the peptide-bound structure has been submitted without this residue being modelled. It is likely that higher-resolution X-ray crystallography data may provide clarity to the existence of an exosite.

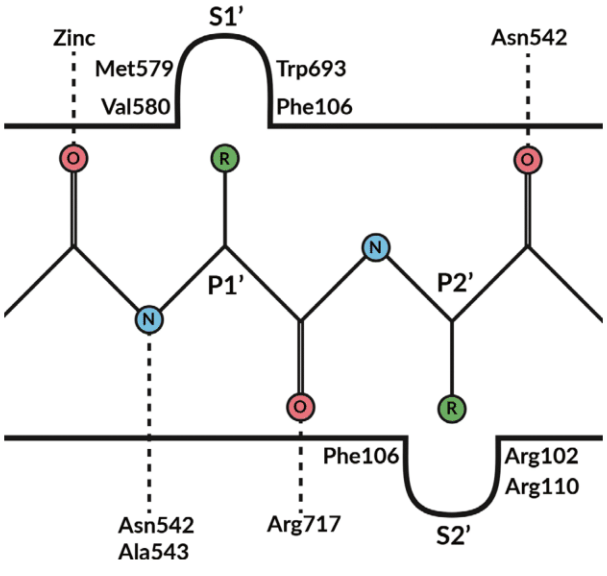


Fig. 7. Schematic of binding interactions between NEP E584D and P1, P1′ and P2′ residues of CNP.

Conclusions

In summary, determination of a peptide-bound NEP E584D has revealed key binding interactions between protein and CNP. The main chain interactions observed for the P1–P3′ residues are likely to be shared for all other NEP substrates.

Given that NEP peptide cleavage can occur at many locations where residue identity is varied, we conclude that peptide-binding interactions between the main chain of P1–P3′ and NEP E584D residues Asn542, Ala543 and Arg717 are most crucial for binding. These important interactions are conserved in cryptidases ECE-1, IDE and PreP. Additional interactions provided by the side chains of P1′ and P2′ residues will contribute to substrate selectivity but may not be as essential to the binding process.

Peptide residues not localised to the active site are likely to exhibit a high degree of flexibility within the central cavity, explaining the inability to identify them in this crystal structure. There was, however, an area of strong electron density observed in a site

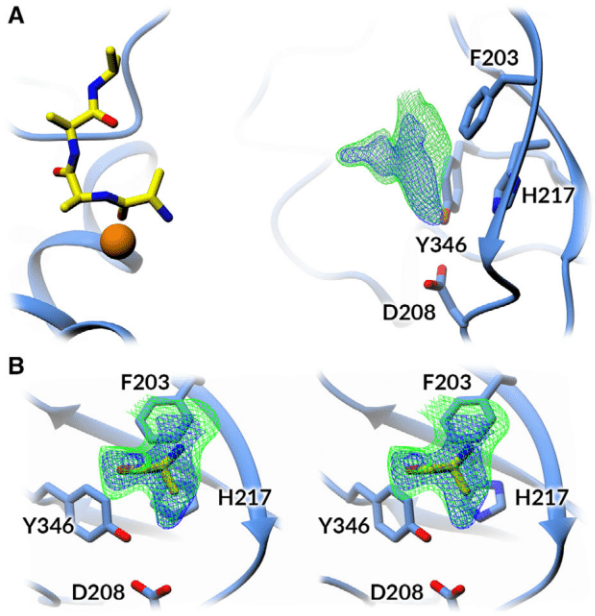


Fig. 8. Putative exosite electron density and nearby residues. (A) $2F_o$-F_c and F_o-F_c omit electron density maps for the putative exosite with close NEP E584D amino acids (6SH2). $2F_o$-F_c and F_o-F_c maps are set to contour levels of 1σ and 3σ, respectively. (B) An alternative orientation of the exosite density with modelled alanine is presented as a stereoview.

proximal to the active site. Similar density in a crystal structure of IDE has been reported as an exosite. Whilst not fully clear, such sites might prove to be critical for peptide binding by conferring the correct positioning of residues within the active site for cleavage. Further higher-resolution crystal structures may allow larger portions of the peptide to be resolved, thereby providing additional information on peptide binding.

Acknowledgements

We thank Diamond Light Source for access to beamline I04 (proposal mx17212) that resulted in the presentation of data here. SM is supported by a postgraduate studentship from the Alzheimer's Society (UK) awarded to KRA and VS [grant number –286 (AS-PhD2015b-006)].

Author contributions

SM performed all the experiments, analysed the data and wrote the manuscript. VS supervised the study and edited the manuscript. KRA conceptualised and supervised the study, analysed the data and edited the manuscript.

Additional information

Accession code: The atomic coordinates and structure factors (codes 6SH1 and 6SH2) have been deposited in the Protein Data Bank (http://www.pdb.org).

References

1 Erdös EG and Skidgel RA (1989) Neutral endopeptidase 24.11 (enkephalinase) and related regulators of peptide hormones. *FASEB J* **3**, 145–151.
2 Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A and Finn RD (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* **46**, D624–D632.
3 Iwata N, Tsubuki S, Takaki Y, Shirotani K, Lu B, Gerard NP, Gerard C, Hama E, Lee HJ and Saido TC (2001) Metabolic regulation of brain Aβ by neprilysin. *Science* **292**, 1550–1552.
4 Bayes-Genis A, Barallat J and Richards AM (2016) A test in context: neprilysin: function, inhibition, and biomarker. *J Am Coll Cardiol* **68**, 639–653.
5 Von Lueder TG, Sangaralingham SJ, Wang BH, Kompa AR, Atar D, Burnett JC and Krum H (2013) Renin-Angiotensin blockade combined with natriuretic peptide system augmentation. *Circ Hear Fail* **6**, 594–605.
6 Korostyshevskaya IM, Maksimov VF and Rudenko NS (2016) C-Type natriuretic peptide: what, where and why? *Neurosci Behav Physiol* **46**, 888–894.
7 Kenny AJ, Bourne A and Ingram J (1993) Hydrolysis of human and pig brain natriuretic peptides, urodilatin, C-type natriuretic peptide and some C-receptor ligands by endopeptidase-24.11. *Biochem J* **291**, 83–88.
8 Xu J, Carretero OA, Liu Y-H, Yang F, Shesely EG, Oja-Tebbe N and Yang X-P (2004) Dual inhibition of ACE and NEP provides greater cardioprotection in mice with heart failure. *J Card Fail* **10**, 83–89.
9 Moss S, Subramanian V and Acharya KR (2018) High resolution crystal structure of substrate-free human neprilysin. *J Struct Biol* **204**, 19–25.
10 Oefner C, D'Arcy A, Hennig M, Winkler FK and Dale GE (2000) Structure of human neutral endopeptidase (neprilysin) complexed with phosphoramidon. *J Mol Biol* **296**, 341–349.
11 Oefner C, Roques BP, Fournie-Zaluski M-C and Dale GE (2004) Structural analysis of neprilysin with various specific and potent inhibitors. *Acta Crystallogr D Biol Crystallogr* **60**, 392–396.
12 Sahli S, Frank B, Schweizer WB, Diederich F, Blum-Kaelin D, Aebi JD, Böhm H, Oefner C and Dale GE (2005) Second-generation inhibitors for the metalloprotease neprilysin based on bicyclic heteroaromatic scaffolds: synthesis, biological activity,

and X-ray crystal-structure analysis. *Helv Chim Acta* **88**, 731–750.
13 Oefner C, Pierau S, Schulz H and Dale GE (2007) Structural studies of a bifunctional inhibitor of neprilysin and DPP-IV. *Acta Crystallogr D Biol Crystallogr* **63**, 975–981.
14 Glossop MS, Bazin RJ, Dack KN, Fox DNA, MacDonald GA, Mills M, Owen DR, Phillips C, Reeves KA, Ringer TJ, *et al.* (2011) Synthesis and evaluation of heteroarylalanine diacids as potent and selective neutral endopeptidase inhibitors. *Bioorg Med Chem Lett* **21**, 3404–3406.
15 Schiering N, D'Arcy A, Villard F, Ramage P, Logel C, Cumin F, Ksander GM, Wiesmann C, Karki RG and Mogi M (2016) Structure of neprilysin in complex with the active metabolite of sacubitril. *Sci Rep* **6**, 27909.
16 Gomis-Rüth F (2008) Structure and mechanism of metallocarboxypeptidases. *Crit Rev Biochem Mol Biol* **43**, 319–345.
17 Schulz H, Dale GE, Karimi-Nejad Y and Oefner C (2009) Structure of human endothelin-converting enzyme I complexed with phosphoramidon. *J Mol Biol* **385**, 178–187.
18 Shen Y, Joachimiak A, Rosner MR and Tang W-J (2006) Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. *Nature* **443**, 870–874.
19 Johnson KA, Bhushan S, Ståhl A, Hallberg BM, Frohn A, Glaser E and Eneqvist T (2006) The closed structure of presequence protease PreP forms a unique 10 000 $Å^3$ chamber for proteolysis. *EMBO J* **25**, 1977–1986.
20 Malito E, Hulse RE and Tang W-J (2008) Amyloid-β degrading cryptidases: insulin degrading enzyme, presequence peptidase, and neprilysin. *Cell Mol Life Sci* **65**, 2574–2585.
21 Schechter I and Berger A (1967) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**, 157–162.
22 Macours N, Poels J, Hens K, Francis C and Huybrechts R (2004) Structure, evolutionary conservation, and functions of angiotensin- and endothelin-converting enzymes. *Int Rev Cytol* **239**, 47–97.
23 Johnson GD and Ahn K (2000) Development of an internally quenched fluorescent substrate selective for endothelin-converting enzyme-1. *Anal Biochem* **286**, 112–118.
24 Waterman DG, Winter G, Gildea RJ, Parkhurst JM, Brewster AS, Sauter NK and Evans G (2016) Diffraction-geometry refinement in the DIALS framework. *Acta Crystallogr D Biol Crystallogr* **39**, 558–575.
25 Evans PR and Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204–1214.
26 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242.
27 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.
28 Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F and Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355–367.
29 Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501.
30 Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS and Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21.
31 Berman H, Henrick K and Nakamura H (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Mol Biol* **10**, 980.
32 McNicholas S, Potterton E, Wilson KS and Noble MEM (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* **67**, 386–394.
33 Teixeira PF, Masuyer G, Pinho CM, Branca RMM, Kmiec B, Wallin C, Wärmländer SKTS, Berntsson RP-A, Ankarcrona M, Gräslund A *et al.* (2018) Mechanism of peptide binding and cleavage by the human mitochondrial peptidase neurolysin. *J Mol Biol* **430**, 348–362.
34 Watanabe Y, Nakajima K, Shimamori Y and Fujimoto Y (1997) Comparison of the hydrolysis of the three types of natriuretic peptides by human kidney neutral endopeptidase 24.11. *Biochem Mol Med* **61**, 47–51.
35 Howell S, Nalbantoglu J and Crine P (1995) Neutral endopeptidase can hydrolyze β-amyloid(1–40) but shows no effect on β-amyloid precursor protein metabolism. *Peptides* **16**, 647–652.
36 Sexton T, Hitchcook LJ, Rodgers DW, Bradley LH and Hersh LB (2012) Active site mutations change the cleavage specificity of neprilysin. *PLoS One* **7**, 1–10.

5.6 Post-paper conclusions

In this chapter the process of determining the first peptide-bound structure of NEP has been presented. Initial attempts to crystallise wild-type NEP with substrate were unsuccessful most likely as a result of substrate cleavage. To overcome the issues surrounding substrate cleavage a mutant NEP E584D protein was generated. Through circular dichroism and X-ray crystallography it was possible to confirm that this the structure of NEP E584D was highly similar to wild-type NEP and therefore a suitable model to study peptide-binding interactions.

Using the mutant protein, a peptide-bound structure of NEP E584D was determined in complex with C-type natriuretic peptide (CNP). This structure revealed the mode of peptide binding for CNP and by extension the likely mode of peptide binding for all substrate. Three key residues Asp 542, Ala 543 and Arg 717 were the main contributors to backbone peptide binding. Comparisons to other zinc metalloproteases revealed the conserved identity of these residues and the interactions they make.

While it was possible to model the backbone of CNP there was limited information available on the side chains of the peptide. This may be a result of several factors, but a likely explanation is that across a protein crystal, individual protein molecules will be binding the CNP at different positions. As such, the side chain density would be a mix of the residues that occupy the subsites. Another obvious limitation with the structure was the resolution. With higher resolution data it may be possible to gain a better indication of side chain identity. Higher resolution data would also provide greater certainty for peptide interactions to areas of the cavity proximal the active site. In the attached manuscript a potential exosite was discussed however at 2.6 Å there is some uncertainty in the density for this site.

Attempts were made to collect high resolution data from a peptide-bound crystal however two issues limited this. The first was the low frequency at which NEP crystals diffracted to high-resolution, an issue discussed in Chapter 3. The second was the low frequency at which crystals either soaked or cocrystallised with substrate showed strong electron density for a bound substrate. Indeed, several datasets for NEP crystals grown or soaked with substrate were collected to a higher resolution but binding of the peptide was not observed in the electron density maps.

Chapter 6 – Construction, expression and purification of endothelin-converting enzyme I

6.1 Foreword

This chapter describes the methods used to express and purify endothelin-converting enzyme I with the intention to use the produced protein for structural and functional characterisation. This results chapter is formatted as a conventional thesis chapter and is not based around a published manuscript.

6.2 Introduction

6.2.1 ECE-1 current structural perspectives

Endothelin-converting enzyme I (ECE-1, EC 3.4.24.71) is a transmembrane zinc metalloprotease. Four isoforms of ECE-1 have been reported - aECE-1, bECE-1, cECE-1 and dECE-1. All isoforms are transcribed from the same gene, however utilisation of different promoters leads to variation in N-terminal size. This variation is minor with the smallest isoform, cECE-1, 754 amino acids in length, while the largest isoform, bECE-1, is 770 amino acids in length. Research has revealed that it is this N-terminal sequence that determines protein localisation. aECE-1 is localised to the plasma membrane, both dECE-1 and cECE-1 are localised to the plasma membrane and intracellular compartments, and bECE-1 is localised almost exclusively intracellularly (Valdenaire *et al.,* 1995; Schweizer *et al.,* 1997; Valdenaire *et al.,* 1999).

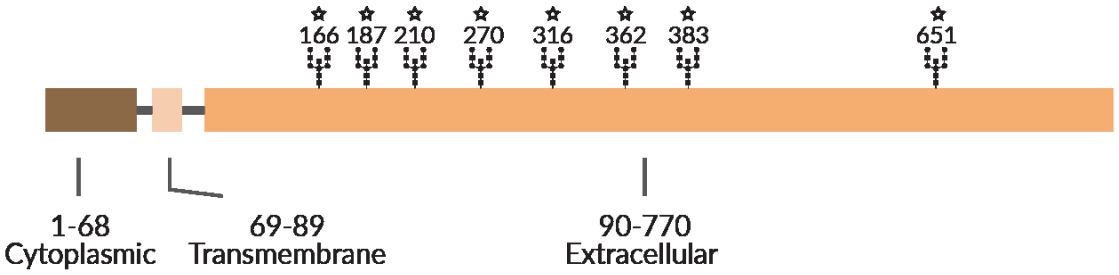
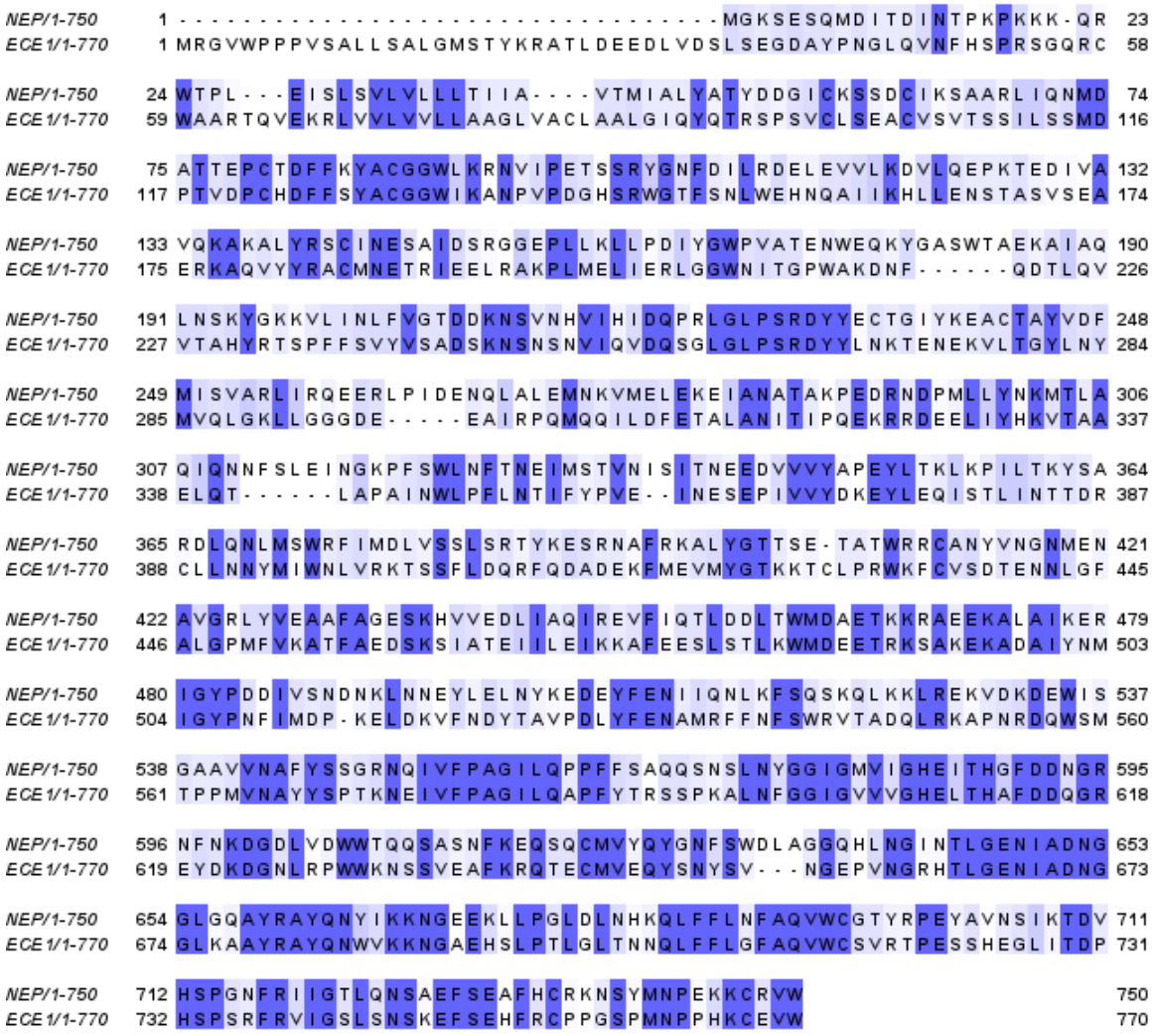


Figure 6.1, ECE-1 domain organisation. ECE-1 is composed of three domains with 8 predicted N-linked glycosylation sites within the extracellular domain (ExPASy).

ECE-1 is composed of three structural domains. For bECE-1, short intracellular and transmembrane domains contain 68 and 21 amino acids respectively, while the large catalytic extracellular domain contains 681 amino acids (Figure 6.1). Human ECE-1 forms a dimer through the formation of a disulfide bond between the Cys428 residue of two molecules. This multimeric organisation has been shown to be important, but not essential,

for ECE-1 activity (Korth *et al.,* 1997). In humans, ECE-1 shares 38% sequence identity with neprilysin (NEP) (Figure 6.2). This figure increases to 53% when considering the extracellular domain alone. Because of this high sequence homology, it is unsurprising that the enzymes have similar roles and specificity. Like NEP, ECE-1 can degrade a wide variety of small peptide substrates that are uniformly < 5 kDa in size (Johnson *et al.,* 1999). While both enzymes have overlapping substrate specificities, they exhibit different cleavage efficiencies. This is most notable in the minimal activity ECE-1 displays for small peptides (< 7 residues) (Johnson *et al.,* 1999).

ECE-1 was initially characterised for its role in conversion of big endothelins to vasoactive endothelins. The production of endothelin-1, a potent vasoconstrictor, has resulted in considerable interest in ECE-1 for its role in blood pressure regulation and cardiovascular disease. As such, ECE-1, like NEP, has been targeted for selective inhibition (Doggrell, 2004). The role of ECE-1 in cardiovascular disease is not limited to its protease activity on big endothelins. Several other vasoactive substrates including natriuretic peptides and bradykinin, can also be degraded by ECE-1 thus providing a further link to blood pressure regulation and cardiovascular disease. It should be noted, however, that unlike the role of ECE-1 in big endothelin degradation, the evidence for protease activity on these substrates *in vivo* is unclear and the physiological relevance has yet to be confirmed (Hartopo *et al.,* 2013; Nakayama *et al.,* 2012).


NEP/1-750 1 - - -

their differences in pH optima, similar levels of activity between have been recorded over a range of substrates (Johnson and Ahn, 2000).

To date, a single structure of the ECE-1 structure has been reported. This crystal structure, containing just the extracellular domain has been determined in an inhibitor-bound state (Schulz *et al.,* 2009). The structure is highly similar to NEP with the domain adopting an ellipsoid shape which can be divided into two lobes connected by a small linker (Figure 6.3a). These lobes act to enclose a large central cavity which contains the active site. The finite volume of this central cavity, like within NEP, has been suggested to be the reason that substrates are limited to <5 kDa.

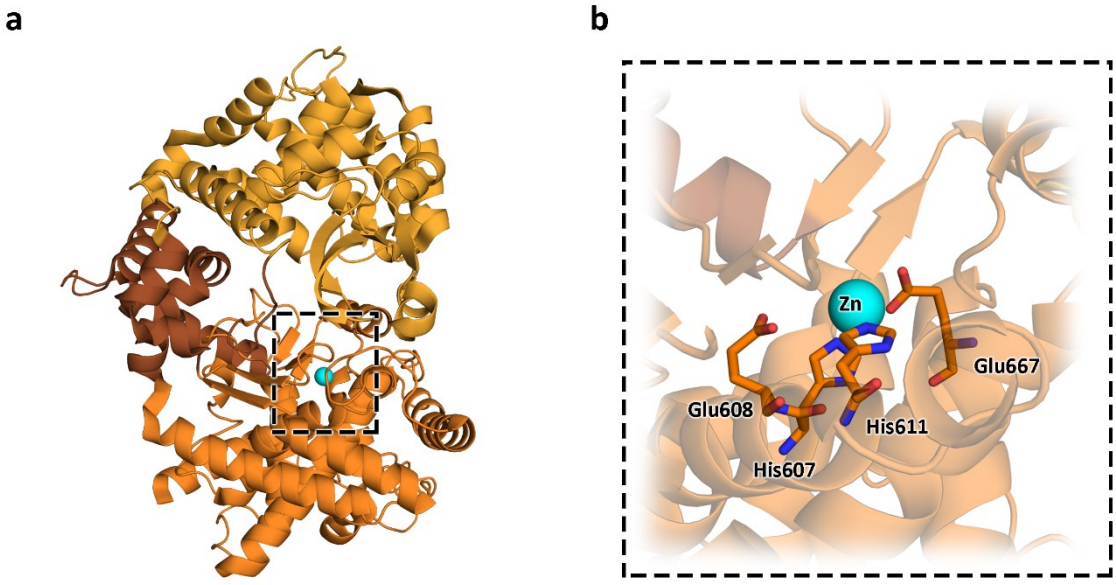


Figure 6.3, Crystal structure of the extracellular domain of ECE-1 (3DWB). (a) ribbon illustration of ECE-1. Subdomains are coloured yellow and orange with the linker region coloured brown. (b) the catalytic site of ECE-1 located around a zinc ion shown in cyan. Here, conserved residues His607 and His611, in addition to Glu667, coordinate the zinc ion. Conserved Glu608 is involved in the proposed catalytic mechanism.

The active site of ECE-1 is based around the conserved HEXXH motif shared by many other zinc metalloproteases. In ECE-1 the residues involved in zinc binding are His607, His611 and Glu667. The conserved Glu608 is positioned above the zinc and is required for the proposed catalytic mechanism. Partial characterisation of the active site has been achieved from the inhibitor-bound structure (Schulz *et al.,* 2009) in addition to site directed mutagenesis studies (Johnson *et al.,* 2002; Schulz *et al.,* 2009). The S1 subsite has very little effect on binding but exhibits some preference for hydrophobic residues. Like NEP, the majority of

ECE-1 specificity is conferred by the S1′ subsite which exhibits a strong preference for large hydrophobic P1′ residues (Rawlings *et al.,* 2018). The S1′ subsite in ECE-1 varies from NEP in just a single residue where Val603 replaces NEP's Met579. While just a single substitution is present there is a substantial increase to the S1′ pocket size.

Considerable information has been gained from the reported inhibitor-bound structure, however, characterisation of substrate binding in ECE-1 has yet to be addressed. Like the work on NEP described in previous chapters, substrate-free and substrate-bound structures of ECE-1 would address several of the underlying questions regarding how substrate binding is achieved and provide greater insights into substrate specificity.

In order to undertake structural and functional characterisation of ECE-1 it is first necessary to establish a method by which the protein can be produced. A method for ECE-1 expression and crystallisations has already been reported (Schulz *et al.,* 2009) and could be used as an initial guideline to determine substrate-free and substrate-bound structures of ECE-1, however, there are several points that need to be noted. First, the inhibitor-bound structure was determined using protein expressed in a bacterial expression system where no mechanisms for glycosylation are present. The protein expressed in this system formed inclusion bodies, likely due to the simplistic protein processing pathway in bacteria. Because of this, protein had to be refolded after expression. The high sequence similarity between ECE-1 and NEP, in conjunction with the high structural similarity between the reported ECE-1 and NEP structures, provides substantial validation to this method of ECE-1 production. Despite this, the absence of information relating to protein glycosylation should be noted, particularly as there was an absence of kinetic data in the published paper.

The second point to note is that the reported inhibitor-bound structure of ECE-1 is in a monomeric state due to point mutation C428S preventing dimer formation. There were attempts to crystallise the wild-type ECE-1, but these were unsuccessful, hence the necessitated for the mutation C428S.

While it would no doubt be preferential to investigate substrate binding in a glycosylated and dimeric ECE-1 protein complex, the reported bacterial expression system and crystallisation protocol provide an attractive starting point. It would, however, be necessary to assess if protease activity is present in the absence of glycosylation and dimerization. If

protease activity is confirmed, structural characterisation could begin. If the protein cannot be refolded or if activity is absent, then alternative expression systems, including yeast and mammalian, could be trialled. If a suitable system is found, active protein could then be used for structural characterisation of substrate binding interactions in ECE-1.

6.3 Materials and Methods

All reagents for expression and purification were sourced from Sigma-Aldrich or Fisher Scientific unless otherwise specified. General methods were followed as described in Chapter 3 (3.2.1).

6.3.1 Buffers

E. coli	
Lysogeny broth (LB) / Lysogeny broth agar	0.5% (w/v) yeast extract, 1% (w/v) peptone and 1% (w/v) NaCl (for LB agar 1% (w/v) agar was added)
Inclusion body solubilisation solution	100 mM Trizma base, 8 M Urea, 10 mM β-mercaptoethanol, 10 mM dithiothreitol (DTT), 1 mM glutathione (GSH), 0.1 mM glutathione disulfide (GSSG) and 1 mM glycine (pH 10, adjusted with NaOH)
Inclusion body wash buffer 1	50 mM Trizma base, 20 mM ethylenediaminetetraacetic acid (EDTA) and 2% (v/v) triton X-100 (pH 7.2, pH adjusted with HCl)
Inclusion body wash buffer 2	50 mM Trizma base and 20 mM EDTA (pH 7.2, pH adjusted with HCl)
E. coli binding buffer 1	50 mM Trizma base (pH 8, pH adjusted with HCl)
E. coli elution buffer 1	50 mM Trizma base and 1 M NaCl (pH 8, pH adjusted with HCl)
E. coli binding buffer 2	50 mM Trizma base and 150 mM NaCl (pH 8, pH adjusted with HCl)
E. coli elution buffer 2	50 mM Trizma base, 150 mM NaCl and 500 mM imidazole (pH 8, pH adjusted with HCl)
P. pastoris	

Yeast Extract Peptone agar (YPD agar)	1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose and 2% (w/v) agar
1 M potassium phosphate buffer, pH 6.0	132 mL of 1 M K_2HPO_4, 868 mL of 1 M KH_2PO_4 (pH 6.0 adjusted with phosphoric acid or KOH)
Buffered complex medium containing glycerol (BMGY)	1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate buffer (pH 6.0), 1.34% (w/v) yeast nitrogen base, 4 × 10^{-5}% (w/v) biotin and 1% (v/v) glycerol. 1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate buffer (pH 6.0), 1.34% (w/v) yeast nitrogen base, 4 × 10^{-5}% (w/v) biotin and 1% (v/v) glycerol. Yeast nitrogen base, biotin and glycerol were filter sterilised and added to the other components which had been autoclaved.
Buffered complex medium containing methanol (BMMY)	1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate buffer (pH 6.0), 1.34% (w/v) yeast nitrogen base, 4 × 10^{-5}% (w/v) biotin and 0.5% (v/v) methanol. 1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate buffer (pH 6.0), 1.34% (w/v) yeast nitrogen base, 4 × 10^{-5}% (w/v) biotin and 1% (v/v) glycerol. Yeast nitrogen base, biotin and methanol were filter sterilised and added to the other components which had been autoclaved.
P. pastoris binding buffer	50 mM Trizma base and 150 mM NaCl (pH 8, pH adjusted with HCl)
P. pastoris elution buffer	50 mM Trizma base, 150 mM NaCl and 500 mM imidazole-HCl (pH 8, pH adjusted with HCl)
HEK 293T	
Dulbecco's Modified Eagle Medium (DMEM)	Sigma-Aldrich (D5796-500ML)
Fetal bovine serum (FBS)	Sigma-Aldrich (F9665-500ML)

0.25 % (w/v) trypsin-EDTA solution	Sigma-Aldrich (T3924-100ML)
Polyethyleneimine, 25k branched	Sigma-Aldrich (12007559)
Phosphate buffered saline solution (PBS)	Sigma-Aldrich (D8537-500ML)
HEK binding buffer	50 mM Trizma base and 150 mM NaCl (pH 8, pH adjusted with HCl)
HEK elution buffer	50 mM Trizma base, 150 mM NaCl and 500 mM imidazole (pH 8, pH adjusted with HCl)
SDS-PAGE buffers	
20x Tris-MOPS running buffer	1.2 M Trizma base, 0.6 M MOPS, 2.0% (w/v) sodium dodecyl sulfate
4x SDS-PAGE sample buffer	250 mM Trizma base, 10% (w/v) sodium dodecyl sulfate, 0.008% (w/v) bromophenol blue, 0.588M β-mercaptoethanol, 40% (v/v) glycerol (pH 6.8, pH adjusted with HCl)
Other buffers	
Enzyme assay buffer	25 mM Bis-Tris-HCl, 150 mM NaCl, 2 mM $MgCl_2$ (pH 6, pH adjusted with HCl)

Table 6.1, Buffers and growth media used for ECE-1 protein expression trials.

6.3.2 Insoluble ECE-1 expression in E. coli

cDNA for the extracellular domain of human ECE-1 (Q90-W770) was purchased from Imagene. The cDNA contained point mutation C428S which has been shown to prevent homodimer formation. The ECE-1 gene was cloned into a pET22b vector and transformed into Rosetta Gami B DE3 (RGB) cells (Novagen) by Dr Mohd Akif from the Acharya lab. Glycerol stocks of the clone were stored at -80°C.

The ECE-1 RGB cell glycerol stock was used to inoculate an LB agar plate containing 100 µg/mL ampicillin, 25 µg/mL chloramphenicol, 50 µg/mL kanamycin, 10 µg/mL tetracycline,

which was incubated for 48 hours at 37°C to establish colonies. A single colony was used to inoculate a starter culture containing 25 mL of LB broth with 100 µg/mL ampicillin, 25 µg/mL chloramphenicol, 50 µg/mL kanamycin and 10 µg/mL tetracycline in a 250 mL flask. The starter culture was incubated overnight at 37°C and 225rpm. 10 mL of the overnight culture was used to inoculate two main cultures in 2 L flasks containing 750 mL LB broth with 100 µg/mL ampicillin and 25 µg/mL chloramphenicol. These main cultures were incubated at 37°C 225rpm until reaching an $OD_{600} \approx 0.6$. The cultures were then induced using IPTG at a final concentration of 2 mM. Main cultures were incubated overnight at 30°C 225rpm. Cultures were transferred to 500 mL centrifuge tubes which were spun at 6000 RCF for 20 minutes at 19°C. The supernatant was discarded, and the cell pellet was used for purification immediately or was flash froze in liquid nitrogen for storage at -80°C.

The ECE-1 RGB cell pellet was resuspended in *E. coli binding buffer 1* with 1 µL benzonase. The resuspended pellet was then passed through a cell disruptor (Constant Systems) pre-cooled to 4°C and washed with purification *E. coli binding buffer 1*. Lysed cells were centrifuged at 47750 RCF for 15 minutes. The supernatant was discarded, and pellet was resuspended in *Inclusion body wash buffer 1*. The resuspended solution was centrifuged at 47750 RCF for 15 minutes. The supernatant was discarded, and pellet was resuspended in *Inclusion body wash buffer 2*. The resuspended solution was centrifuged at 47750 RCF for 15 minutes. The supernatant was discarded and the inclusion body containing pellet was solubilised in *inclusion body solubilisation solution* at 0.1 g/mL and stirred at room temperature for 1-hour. Insoluble matter was separated by centrifugation at 47750 RCF for 15 minutes. The solubilised fraction was diluted to an OD_{280} of 5 in *inclusion body solubilisation solution* and used in a refolding screen.

For the refolding screen 50 µL of solublised sample was added to 950 µL of refolding conditions (Table 6.2) and incubated at 4°C. After 24 hours samples were assessed using SDS-PAGE. Large-scale refolding was achieved using 450 mL of *refolding buffer 13* (Table 6.2) and 50 mL of solubilised inclusion body. Refolding was performed for 24 hours at 4°C after which the protein solution was loaded at 5 mL/min onto a 5 mL HiTrap anion exchange column (GE Healthcare) pre-eqilibrated with *E. coli binding buffer 1*. Protein elution was achieved using a 10% stepwise elution of *E. coli elution buffer 1*.

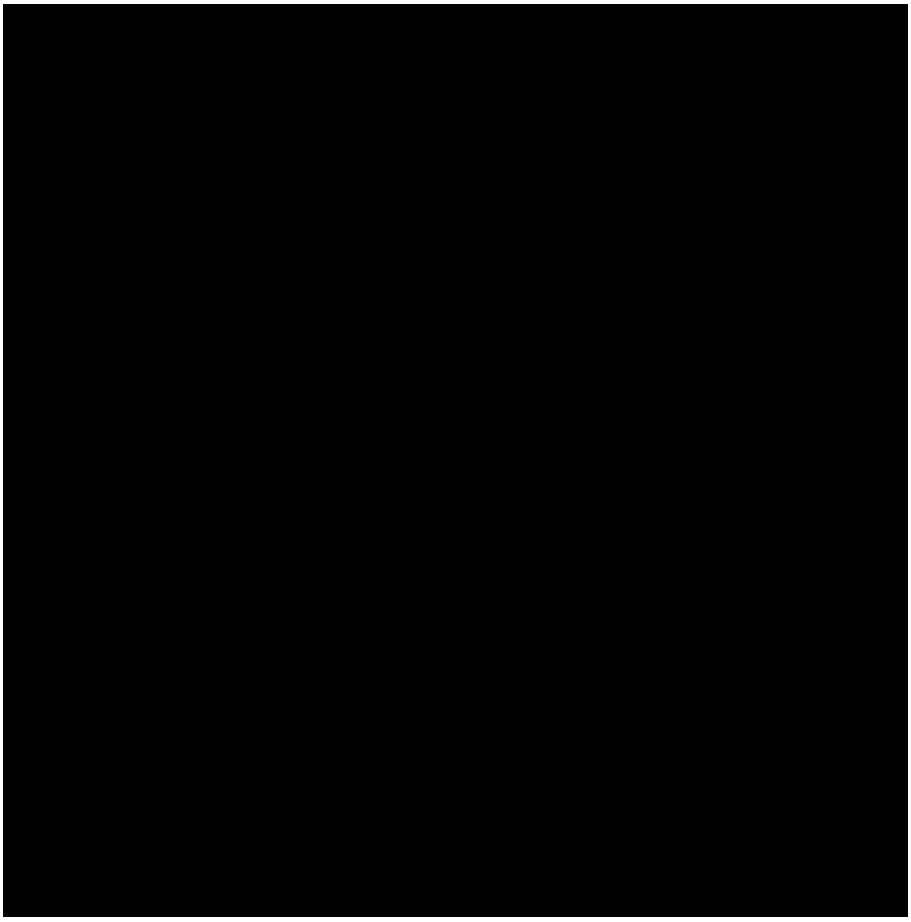


Table 6.2, ECE-1 in-house protein refolding screen. 16 conditions with varied pH and chemical components were used to refold ECE-1 protein from inclusion bodies.

6.3.3 Soluble ECE-1 expression in E. coli

Human ECE-1 C428S pET22b cDNA was provided by Dr Mohd Akif. This cDNA was used as a template for a polymerase chain reaction (PCR) in which the entire vector was amplified using primers listed in Table 6.3. PCR reaction components and parameters are detailed in Table 6.4 and Table 6.5. This PCR reaction added a 5′ 6x histidine tag upstream of the ECE-1 gene for western blot analysis and ease of protein purification.

Primer	**DNA sequence**
Amplification forward	GAAGGAGATATACATATGCATCACCACCATCATCACCAGTACCAG ACAAGATCCCCCTCTG
Amplification reverse	CAGAGGGGGATCTTGTCTGGTACTGGTGATGATGGTGGTGATGC ATATGTATATCTCCTTC

Table 6.3, Primer DNA sequences used to generate histidine tagged ECE-1 C428S pET22b.

Component	Reaction Volumes	Final Concentration
5X Q5 Reaction Buffer	10 µL	1X
10 mM dNTPs	1 µL	200 µM
10 µM Forward Primer	2.5 µL	0.5 µM
10 µM Reverse Primer	2.5 µL	0.5 µM
1 ng/µL Template DNA	1 µL	0.02 ng/µL
Q5 Hot Start High-Fidelity DNA Polymerase	0.5 µL	0.02 U/µL
Nuclease-Free Water	32.5 µL	

Table 6.4, Standard PCR reaction components.

Stage	Temperature	Time
Initial Denaturation	98 °C	30 seconds
35 Cycles ↓	98 °C	10 seconds
	60 °C	20 seconds
	72 °C	4 minutes
Final Extension	72 °C	10 minutes
Hold	4 °C	

Table 6.5, PCR reaction parameters used to generate histidine tagged ECE-1 C428S pET22b.

The amplified histidine tagged ECE-1 C428S pET22b was transformed into RGB cells. Transformation was performed by incubating 20 µL RGB cells on ice with 1 µL of 1 ng/µL DNA for 10 minutes. Cells were transferred to a water bath at 42°C for 45 seconds before incubating on ice for a further two minutes. 100 µL of LB was added to the cells and cells were incubated at 37°C for 45 minutes at 200 rpm. 100 µL cell solution was then plated onto an LB agar plate containing 100 µg/mL ampicillin, 25 µg/mL chloramphenicol, 50 µg/mL kanamycin, 10 µg/mL tetracycline. LB agar plates were incubated overnight at 37°C.

A starter culture containing 25 mL of LB broth with 100 µg/mL ampicillin, 25 µg/mL chloramphenicol, 50 µg/mL kanamycin and 10 µg/mL tetracycline in a 250 mL flask was inoculated with a single ECE-1 C428S pET22b RGB colony and incubated overnight at 37°C

and 225rpm. 1 mL of the overnight culture was used to inoculate test expression cultures in 250 mL flasks containing 75 mL LB broth with 100 µg/mL ampicillin and 25 µg/mL chloramphenicol. These main cultures were incubated at 37°C 225rpm until reaching an OD_{600} ≈ 0.6. The cultures were then induced using IPTG at a final concentration of 2 mM. Induction times and temperatures were varied to determine an appropriate condition for soluble protein expression. Samples of the test expressions were analysed with SDS-PAGE and western blot.

Once an appropriate condition for soluble ECE-1 expression was determined large-scale protein expressions were performed. A starter culture containing 25 mL of LB broth with 100 µg/mL ampicillin, 25 µg/mL chloramphenicol, 50 µg/mL kanamycin and 10 µg/mL tetracycline in a 250 mL flask was inoculated with a single ECE-1 C428S pET22b RGB colony and incubated overnight at 37°C and 225rpm. 10 mL of the overnight culture was used to inoculate two main cultures in 2 L flasks containing 750 mL LB broth. These main cultures were incubated at 37°C 225rpm until reaching an OD_{600} ≈ 0.6. Induction was achieved using IPTG at a final concentration of 2 mM. A 3-hour induction at 16°C was selected for soluble protein expression. After this induction period, cells were harvested from expression cultures by centrifugation for 20 minutes at 5000 RCF. The supernatant was discarded, and cell pellet was resuspended in *E. coli binding buffer 2* (20 mL per g of cell pellet) with 1 µL benzonase. The resuspended pellet was then passed through a cell disruptor (Constant Systems) pre-cooled to 4°C and washed with *E. coli binding buffer 2*.

The soluble and insoluble cell lysate fractions were separated by centrifugation for 30 minutes at 47750 RCF. The supernatant was collected and filtered through a 0.22 µm syringe filter (Millipore). Filtered supernatant was loaded at 5 mL/min onto a 5 mL histidine affinity column (GE Healthcare) pre-equilibrated with *E. coli binding buffer 2*. Protein elution was achieved using a gradient of *E. coli elution buffer 2* over 40 mL. Samples were taken from fractions with slightly elevated UV absorbances and analysed with SDS-PAGE and western blot.

6.3.4 ECE-1 expression in Pichia pastoris

Human ECE-1 C428S cDNA with a 5′ histidine tag was amplified using PCR from the previously created ECE-1 C428S pET22b. Primers for this PCR reaction were designed to add

a 5′ EcoRI site and a 3′ NotI site to the ECE-1 gene for cloning purposes (Table 6.6). The PCR reaction components and parameters are detailed in Table 6.6 and Table 6.7.

The amplified product and vector pPICzαA were digested with restriction enzymes EcoRI and NotI (NEB) (Table 6.8). Digested DNA was then ligated with T4 ligase (NEB) to form ECE-1 C428S pPICzαA (Table 6.9).

Primer	**DNA sequence**
Amplification forward	CCGCCGGAATTCCATCACCACCATCATCACGGAGG
Amplification reverse	GCTTGCGGCCGCTTACCAGACTTCGCACTTGTGAGGCG

Table 6.6, Primer DNA sequences used to generate histidine tagged ECE-1 C428S for ligation into pPICzαA.

Stage	**Temperature**	**Time**
Initial Denaturation	98 °C	30 seconds
35 Cycles ↓	98 °C	10 seconds
	60 °C	20 seconds
	72 °C	90 seconds
Final Extension	72 °C	10 minutes
Hold	4 °C	

Table 6.7, PCR reaction parameters used to generate histidine tagged ECE-1 C428S for ligation into pPICzαA.

Component	**Reaction Volumes**
Restriction Enzyme	1 µL (10U) per enzyme
DNA	1 µg
10X NEBuffer	5 µL (1X)
Nuclease-free water	to 50 µL
Incubation Time	1 hour
Incubation Temperature	37 °C

Table 6.8, Standard reaction parameters for a restriction digest.

Component	Reaction Volumes
T4 DNA Ligase Buffer (10X)	2 µL
Vector DNA (3 kb)	50 ng
Insert DNA (2.2 kb)	110 ng
T4 DNA Ligase	1 µL
Nuclease-free water	to 20 µL
Incubation Time	16 hours
Incubation Temperature	16 °C

Table 6.9, Ligation reaction parameters used to generate histidine tagged ECE-1 C428S pPICzαA.

The ECE-1 C428S pPICzαA construct was transformed into TOP10 Chemically Competent *E. coli* cells (ThermoFisher) for plasmid replication (protocol detailed in 5.3.1). DNA was harvested from TOP10 cells using Wizard® Plus SV Minipreps DNA Purification Systems (Promega). DNA was linearised with restriction enzyme PmeI (NEB) using standard parameters (Table 6.8) and integrated into the genome of *P. pastoris* GS115 using electroporation in accordance with the Invitrogen EasySelect manual (cat. no. K1740-01). Transformed cells were then plated onto a YPD agar plate containing 100 µg/mL zeocin for colony selection.

Protein expression using transformed *P. pastoris* clones was done in accordance with the Invitrogen EasySelect expression manual (cat. no. K1740-01). A single colony was used to inoculate a starter culture containing 25 mL of BMGY media in a 250 mL baffled flask. The starter culture was incubated overnight at 30°C 225rpm. 10 mL of the overnight culture was used to inoculate a 250 mL BMGY media in a 2 L baffled flask. The main culture was incubated at 30°C and 225rpm for 24 hours[1].

After 24 hours the main culture was transferred to 500 mL centrifuge tubes which were centrifuged at 6000 RCF for 20 minutes at 19 °C. The supernatant was discarded, and the cell pellet was resuspended in 250 mL of BMMY media for induction. Resuspended cells were transferred to a new 250 mL baffled flask. The culture was then incubated for 24 hours at 30 °C 225rpm.

[1] For small-scale test expressions the same protocol was used but the culture volume was scaled down to 25 mL.

After 24 hours of induction, the main culture was transferred to a 500 mL centrifuge tubes and spun at 6000 RCF for 20 minutes at 19 ºC. The supernatant was collected, and the cell pellet was discarded. Solid Trizma base and NaCl were added to the supernatant to give final concentrations of 25mM Trizma and 150mM NaCl. The supernatant was mixed with magnetic stirrer for 10 minutes at room temperature. After this time, the supernatant showed some precipitation which was removed by flowing through a 0.22 µm filter (Millipore). The final filter supernatant had a pH of ~7.

The buffer equilibrated *P. pastoris* supernatant was loaded onto a 5 mL HisTrap HP affinity column (GE Healthcare Life Sciences) at 2 mL/min pre-equilibrated with *P. pastoris binding buffer*. Following loading, the column was washed with *P. pastoris binding buffer* until the UV trace returned to baseline. Protein elution was performed at 5 mL/min in a stepwise manner. 2% *P. pastoris elution buffer* was used to wash the column until UV trace stabilised. Bound protein was eluted using 50% *P. pastoris elution buffer* and 2 mL fractions were collected.

Protein containing fractions were confirmed using SDS-PAGE gel electrophoresis. These fractions were concentrated to under 2 mL in a centrifuge at 4000 RCF using a 30kDa molecule weight cutoff filter unit (Millipore) The concentrated protein containing sample was then loaded at 1.5 mL/min onto a Superdex 16/60 pre-equilibrated with *P. pastoris binding buffer*.

6.3.5 ECE-1 expression in human embryonic kidney cells

The human ECE-1 C428S gene was amplified using PCR from the previously created ECE-1 C428S pPICzαA. Primers for this PCR reaction were designed to add a 5′ KpnI site and a 3′ PmeI site to the ECE-1 gene for cloning purposes (Table 6.10). The PCR reaction components and parameters are detailed in Table 6.6 and Table 6.11. The amplified product and vector pOPINTTGneo were digested with restriction enzymes KpnI and PmeI (NEB) using the reaction parameters described in Table 6.8. Digested DNA was then ligated with T4 ligase (NEB) to form ECE-1 C428S pOPINTTGneo using reaction parameters detailed in Table 6.12.

Primer	**DNA sequence**
Amplification forward	GTAGCTGAAACCGGGCACCATCATCACCACCATCATCACGG
Amplification reverse	GTGATGGTGATGTTTTTACCAGACTTCGCACTTGTGAGG

Table 6.10, Primer DNA sequences used to generate histidine tagged ECE-1 C428S for ligation into pOPINTTGneo.

Stage	Temperature	Time
Initial Denaturation	98 °C	30 seconds
35 Cycles ↓	98 °C	10 seconds
	58 °C	20 seconds
	72 °C	90 seconds
Final Extension	72 °C	10 minutes
Hold	4 °C	

Table 6.11, PCR reaction parameters used to generate histidine tagged ECE-1 C428S for ligation into pOPINTTGneo.

The ECE-1 C428S pOPINTTGneo construct was transformed into TOP10 Chemically Competent *E. coli* cells (ThermoFisher) for plasmid replication (protocol detailed in 5.3.1). DNA was harvested from TOP10 cells using Wizard® Plus SV Minipreps DNA Purification Systems (Promega). DNA then transformed into HEK 293T cells. A small-scale expression in a 24-well plate was used to determine the best transformation protocol for the highest protein expression.

Component	Reaction Volumes
T4 DNA Ligase Buffer (10X)	2 µL
Vector DNA (3 kb)	50 ng
Insert DNA (2.2 kb)	35 ng
T4 DNA Ligase	1 µL
Nuclease-free water	to 20 µL
Incubation Time	16 hours
Incubation Temperature	16 °C

Table 6.12, Ligation reaction parameters used to generate histidine tagged ECE-1 C428S pOPINTTGneo.

HEK 293T cells were revived from cryogenic storage with a 2-3 minute thaw in a 37°C water bath. 1 mL of thawed cells was added to 9 mL complete growth medium (pre-warmed to 37°C). The cell suspension was centrifuged at 100 RCF for 3 minutes before removing the

supernatant and replacing with 15 mL complete growth medium (pre-warmed to 37°C). The cell suspension was then transferred to a T75 flask and incubated at 37°C in a 5% CO_2 humidified incubator. Cells were passaged when cultures reached between 70 - 90% confluence.

20 - 24 hours before transfection HEK 293T cells were dissociated from a T175 flask and diluted to between 1 - 2 x 10^5 viable cells/mL. Cells were then transferred as 500 µL aliquots to separate wells of a 24-well plate. Cultures were incubated at 37°C in a 5% CO_2 humidified atmosphere until between 40-80% confluent. 1-4 hours before preparing transfection complexes cell growth medium was replaced with 500 µL reduced serum media.

To create transfection complexes 100 µL aliquots of serum free medium were made. To each aliquot 20 - 40 µL of 100 ng/µL DNA and 2 - 16 µL of 1 µg/µL PEI was added. Aliquots were mixed using a vortexer for 20-30 seconds and incubated at room temperature for up to 10 minutes to allow complex formation. Transfection complexes were added dropwise to each 500 µL culture and incubated at 37°C in a 5% CO_2 humidified atmosphere. Cell supernatant was harvested 24 hours after transfection.

Once appropriate transfection conditions had been determined, large-scale transfections were performed in an expanded surface (2125 cm^2) roller bottles (Cellmaster, Greiner Bio-One). Roller bottles were inoculated from cells grown in a T175 flask at >90% confluency. Cells were removed from the T175 flask by first removing cell growth media. Cells were then washed with 10 mL PBS before adding 5 mL of trypsin solution to dissociate attached cells. The resuspended cells in 5 mL trypsin solution was then added to 20 mL DMEM containing 10% (v/v) FBS (Dulbecco's Modified Eagle Medium, ThermoFisher Scientific, 41965039) which was then added to 225 mL DMEM containing 10% (v/v) FBS (pre-warmed to 37 °C) in the roller bottle. Roller bottles were incubated at 37°C in a 5% CO_2 whilst rotating at 0.5 rpm.

After 24 hours, growth media was removed and replaced with 200 mL of DMEM containing 2% (v/v) FBS. To this, the transfection solution – 2 mg of ECE-1 C428S pOPINTTGneo in 50 mL of DMEM containing 2% (v/v) FBS – was added to the roller bottle. A further incubation was then performed at 37°C in a 5% CO_2 whilst rotating at 0.5 rpm. Cell supernatant was harvested 4 days after transfection.

The 250 mL of cell media was removed from roller bottles and centrifuged at 4000 RCF for 30 minutes. Cell media was then filtered through a 0.22 µm syringe filter (Millipore) before loading at 5 mL/min onto a HisTrap HP affinity column (GE Healthcare) pre-equilibrated in *HEK binding buffer*. Protein elution was achieved using a gradient (0 - 100%) of *HEK elution buffer* over 40 mL. Eluted fractions were then analysed using SDS-PAGE and western blot. ECE-1 containing fractions were concentrated and pooled before loading at 1.5 mL/min onto a Superdex 16/60 size exclusion column pre-equilibrated in *HEK binding buffer*.

6.3.6 Cleavage assay

The protease activity of ECE-1 C428S was assessed by measuring cleavage of the fluorogenic peptide Mca-RPPGFSAFK-(Dnp) (Enzo Life Sciences). Protein was diluted to varied concentrations in *Enzyme assay buffer*. 50 µL of diluted protein was added to a black 96 well microplate (cat. no. 655076, Greiner Bio-One). 50 µL of Mca-RPPGFSAFK-(Dnp), between 50 - 3.125 µM dissolved in *Enzyme binding buffer* with 10% (v/v) DMSO, was added to a black 96 well microplate (Greiner Bio-One) to initiate the reaction.

Each fluorescence assay was performed in triplicate where samples allowed and fluorescence was recorded at 20 second intervals on a CLARIOstar high-performance microplate reader (BMG LABTECK). Excitation was achieved at 328 nm and emission was detected at 393 nm. Values for initial rates were determined in relative fluorescence units (RFU) using a line of best fit through linear data points up to 200 seconds.

6.4 Results

6.4.1 Insoluble ECE-1 expression in E. coli

Initial protein expression and refolding protocols followed the methods described in Schulz *et al*. (2009), which had been used to determine the structure of the ECE-1 extracellular domain in an inhibitor-bound state. A Rosetta-gami B (RGB) clone containing the extracellular domain of human ECE-1 (Q90-W770) C428S in vector PET22b was created by the Dr Mohd Akif from the Acharya lab (Figure 6.4). A key benefit of the cell line was the advanced expression of eukaryotic proteins and improved disulfide bond formation.

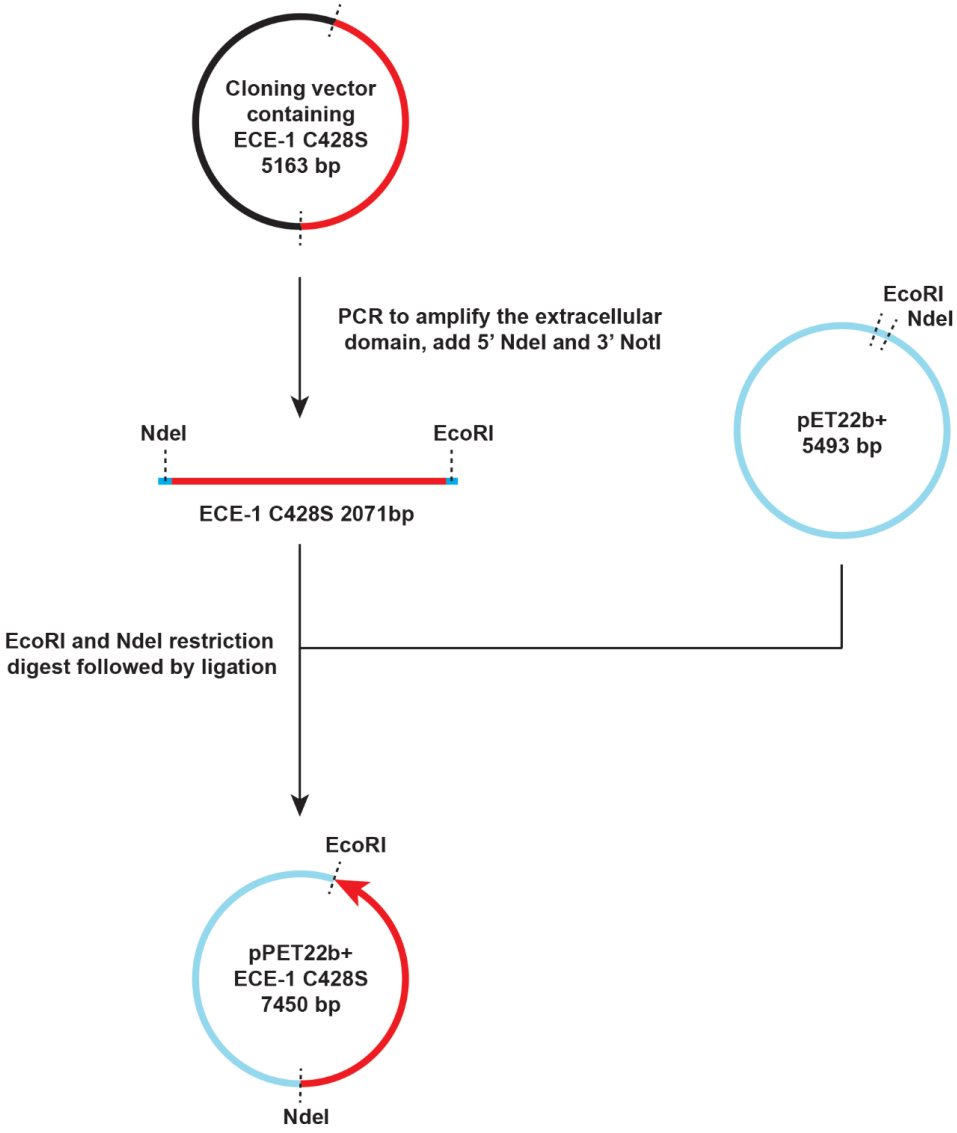


Figure 6.4, Cloning methods conducted for generation ECE-1 C428S pET22b. Cloning was performed by Dr Mohd Akif.

The protein expression protocol followed a 30°C growth and induction temperature. Protein formed during induction was found within insoluble inclusion bodies in the cell lysate. Inclusion bodies were separated from the rest of the cell mass and SDS-PAGE analysis confirmed the presence of an ~80 kDa protein corresponding to ECE-1 (Figure 6.5). SDS-PAGE analysis also revealed the presence of several other molecular weight bands in the insoluble sample, most notably at ~30 and ~45 kDa. From this analysis, it was possible to confirm that ECE-1 expression in RGB cells could yield large quantities of protein, however the insoluble protein formed would require refolding and further purification.

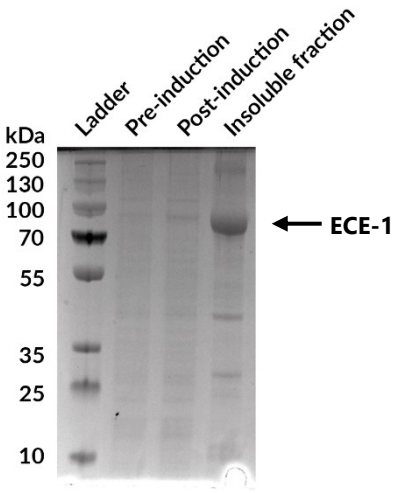


Figure 6.5, ECE-1 expression in RGB cells. Pre- and post- induction cell lysate and the insoluble fraction of the post-induction cell lysate was analysed using SDS-PAGE gel electrophoresis. The band corresponding to ECE-1 is marked with an arrow.

Information regarding protein refolding was not included in the protocol described in Schulz *et al*. (2009). Instead, an in-house refolding screen was performed to determine a condition in which ECE-1 from inclusion bodies could form soluble and correctly folded protein. 16 conditions were screened, and protein samples incubated in those conditions were analysed by SDS-PAGE (Figure 6.6). Six refolding conditions contained protein that was soluble after a 24-hour incubation. Other refolding conditions contained large amounts of precipitated protein. Protein samples from these six conditions were then assessed for their peptide cleavage using the assay described in Chapter 3 and 5.

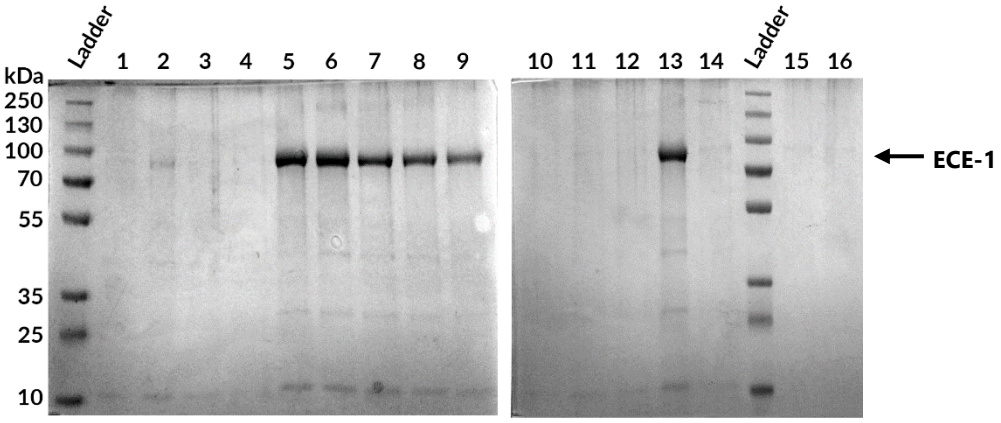


Figure 6.6, ECE-1 after incubated in refolding conditions. ECE-1 expressed in inclusion bodies was solubilised before being added to 16 refolding conditions. After a 24-hour incubation at 4°C, samples were loaded onto an SDS-PAGE

to assess which conditions resulted in soluble ECE-1 protein. The bands corresponding to ECE-1 is marked with an arrow.

Full kinetic characterisation for each ECE-1 refolded samples was not performed. Instead, a single substrate concentration used for each protein sample. While full characterisation would have been preferential, it was not required to provide insights into the efficiency of protein refolding. It has been reported that ECE-1 should possess a 10-fold greater efficiency for substrate Mca-RPPGFSAFK-(Dnp) compared to NEP (Johnson and Ahn, 2000). Therefore, the assay was used to provide a rudimental estimate as to whether refolding had been successful in refolding ECE-1. It is also important to note the assays were conducted using varied protein concentrations. The rationale behind this was to allow accurate measurement of initial rate. In the case of the refolded ECE-1 samples, far higher protein concentrations were required compared to control NEP samples to provide detectable fluorescence.

Two of the six conditions assayed displayed some level of peptide cleavage, however the recorded initial rate was substantially lower than that observed for NEP despite the far greater protein concentrations (Figure 6.7). The greatest initial rate value of the refolded samples was detected in ECE-1 incubated in refolding condition 13. The refolding protocol for this condition was repeated on a larger scale and refolded protein was used to determine the suitability of the sample for protein purification. Refolded protein was purified using anion exchange in accordance with the protocol described in (Schulz *et al.,* 2009). A gradient of NaCl containing buffer was used to elute the protein in a stepwise manner and several peaks were observed (Figure 6.8).

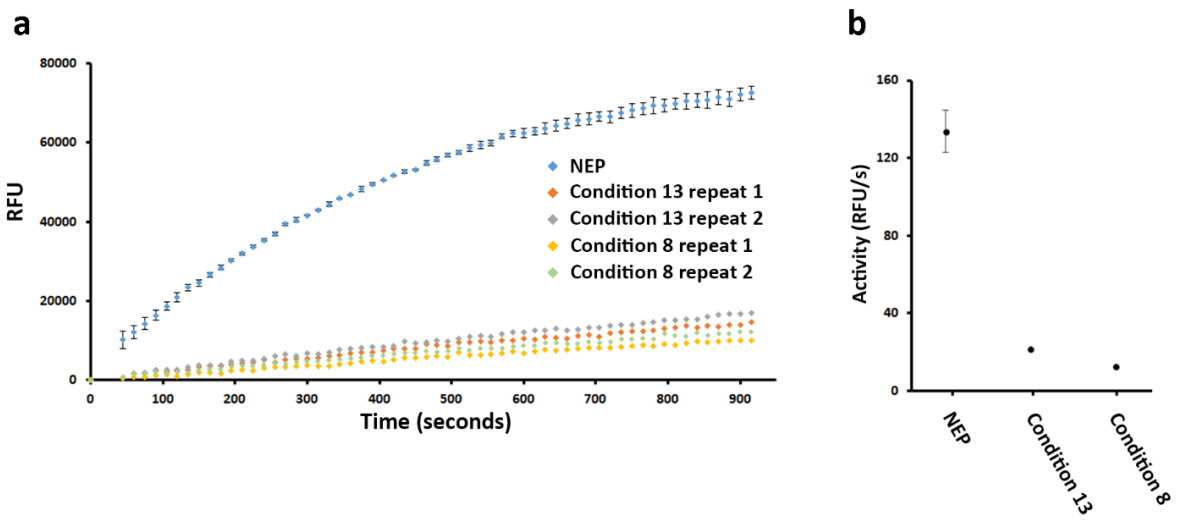


Figure 6.7, Peptide cleavage assay for ECE-1 after refolding. (a) Peptide cleavage was assessed for 5 µM ECE-1 refolded in conditions 8 and 13 at a single substrate concentration of 10 µM. NEP was used as a positive control at 4 nM. Substrate cleavage by NEP was repeated three times with the mean and ±SD error bars shown above. For ECE-1 samples, substrate cleavage was repeated twice with both repeats shown above. (b) Initial rates calculated from the peptide cleavage assays. For NEP, the mean and ±SD error bars are shown from the initial rates calculate from three repeats. For ECE-1, the initial rate is shown as the mean of two repeats.

Eluted protein fractions corresponding to peaks were analysed using SDS-PAGE (Figure 6.9) to reveal all eluted peaks contained a band corresponding to ECE-1 (~80 kDa). The most likely explanation for this was that protein in the refolding buffer was present in a range of partially folded protein molecules. These differently folded species had different surface exposed regions and therefore bound with varied affinity to the anion exchange column. Refolding was repeated and attempted for longer time periods however similar results were observed.

The generation of heterogenous species was highly detrimental to protein yield and was inappropriate for protein crystallisation. It may have been possible, through screening more conditions and incubation times, to improve the suitability of refolded protein for structural characterisation. However, initial results for protein refolding did not indicate that it was an efficient use of time. Instead focus was placed on expressing soluble and correctly folded protein.

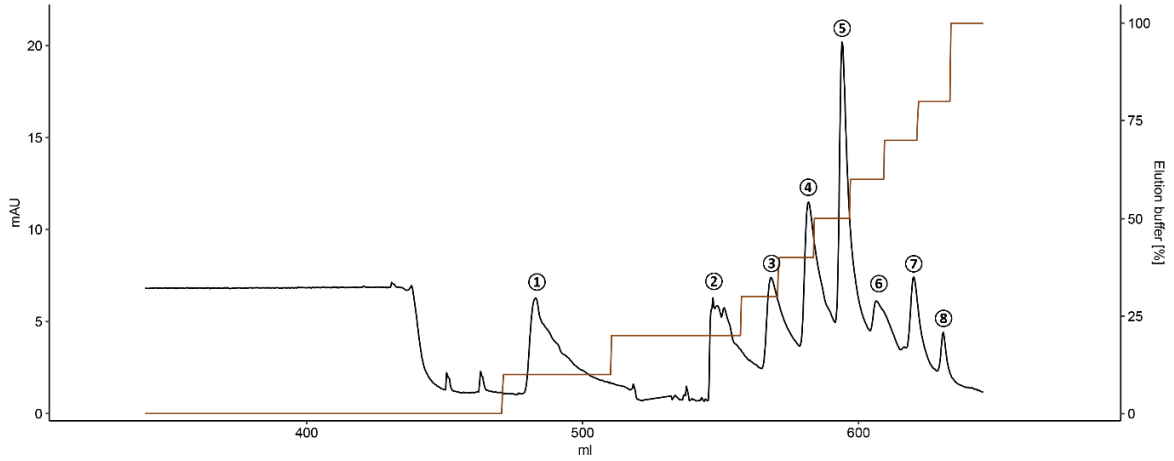


Figure 6.8, Anion exchange purification of refolded ECE-1. 500 mL of protein refolded in refolding condition 13 was loaded onto an anion exchange column (GE healthcare). Once loaded the column was used with E. coli loading buffer 1 until the UV trace returned to baseline. Protein elution was achieved using 10% step increases in E. coli elution buffer 1. A total of eight peaks were observed after step increases in elution buffer. Fractions of these peaks were analysed with SDS-PAGE.

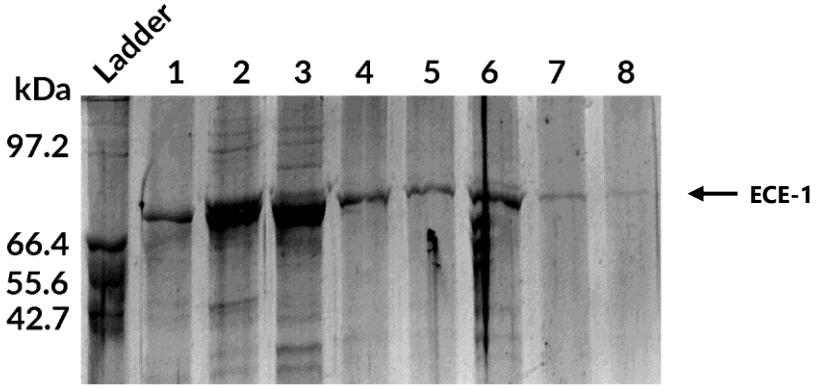


Figure 6.9, ECE-1 anion exchange protein containing fractions. Sample were taken from eight peak containing fractions from the anion exchange purification and analysed with SDS-PAGE electrophoresis. All samples contained a protein band at ~80 kDa corresponding to ECE-1 which is marked with an arrow.

For soluble ECE-1 expression three expression systems were selected and tested in parallel. The first system, *P. pastoris*, was an obvious choice based on the previous success seen in NEP expression. The second system, human embryonic kidney 293T cells, was selected for its mammalian post-translational machinery. Finally, in addition to the two new expression systems, the RGB expression system was revisited to determine if it was possible to express soluble ECE-1 through variation in induction conditions.

6.4.2 Soluble ECE-1 expression in E. coli

Expression of ECE-1 in RGB *E. coli* cells was revisited to determine if soluble protein could be formed. The protocol described in Schulz *et al*. (2009) used a 30°C growth and induction temperature to generate ECE-1 that formed inclusion bodies. At high temperatures such as this has been reported to greatly increase the likelihood of inclusion body formation due to the increased rate of protein production in conjunction with the limited capacity of protein folding pathways (Singh *et al.,* 2010). To determine if it was possible to produce soluble protein in RGB cells, several expression conditions with varied induction times and temperatures were assessed. Prior to this expression screening the initial DNA construct was altered such that the ECE-1 gene contained a 5′ 6x histidine tag. This tag allowed western blot analysis using poly histidine antibodies and had the added benefit of improving purification efficiency. A whole plasmid PCR reaction was used to generate the final 6x histidine tag containing construct (Figure 6.10). This final vector was analysed with an agarose gel (Figure 6.11), sequenced and transformed into RGB cells for expression.

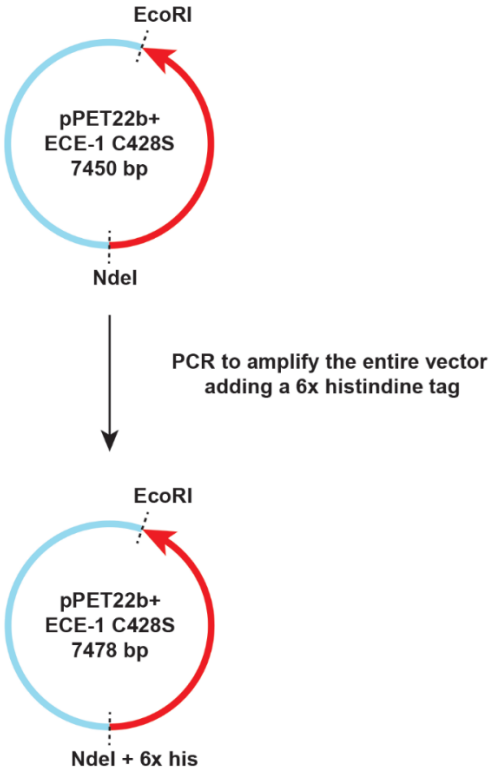


Figure 6.10, Cloning plan to modify the pET22b ECE-1 C428S construct by adding a 5′ 6x histidine tag. The tag allowed western blot analysis as well as more efficient purification using a histidine trap affinity column.

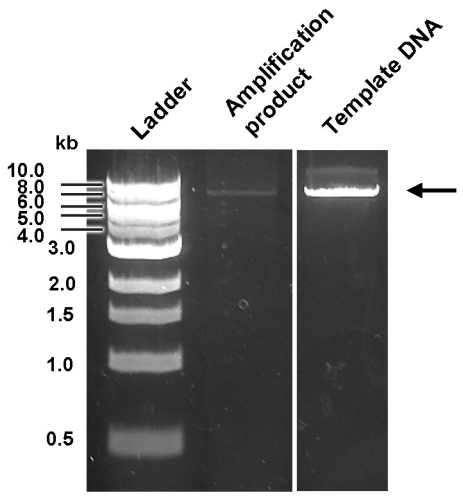


Figure 6.11, Generation of ECE-1 pET22b C428S with a 5′ 6x histidine tag. The PCR amplification product was analysed using agarose gel electrophoresis revealing a single band is present at ~7.5 kDa (marked with an arrow). The band for the template DNA is has been included as a size reference.

Six expression protocols were tested for soluble ECE-1 expression (Table 6.13). Samples of the soluble and insoluble supernatant and cell lysate fractions were collected before and after induction for each expression condition. These samples were then analysed with SDS-PAGE and western blot (Figure 6.12).

Analysis of induction temperatures revealed that a large amount of protein formed inclusion bodies when induction was performed at 37°C or 20°C for 20 hours (Figure 6.12b). In these conditions a large band at ~80 kDa corresponded to the extracellular domain of ECE-1, however, several other lower molecule weight bands were also present on the western blots. These additional bands likely correspond to degradation products of ECE-1 which occurred to a greater extent with increased temperature and induction length.

A single induction condition - '3 hours at 16°C' - contained a band corresponding to ECE-1 in the soluble cell lysate sample (Figure 6.12b). This condition emerged as the most promising of all conditions for generating soluble ECE-1 protein. A second protein band at ~45 kDa was also present in this sample and likely corresponded to degradation product.

Based on the results observed in the trial expressions the '3 hours at 16 °C' condition was chosen for large-scale protein production. The soluble cell lysate from a large-scale expression was loaded onto a histidine trap affinity column for purification (Figure 6.13). Protein elution was achieved using a gradient of imidazole containing buffer. Eluted samples

were then analysed using SDS-PAGE. From this analysis it was clear that while ECE-1 protein was present in the loaded sample (Figure 6.14b), it could not be identified in the column eluate. Despite repeated expression and purification, the same outcome was observed. As only low level ECE-1 was present in the loaded sample it is likely that the sample became too dilute and could not be detected in the eluate.

Sample number	Condition	
1	Pre-induction	S/N
2		Cell lysate soluble
3		Cell lysate Insoluble
4	Post-induction 3hrs at 4°C	S/N
5		Cell lysate soluble
6		Cell lysate Insoluble
7	Post-induction 20hrs at 4°C	S/N
8		Cell lysate soluble
9		Cell lysate Insoluble
10	Post-induction 3hrs at 16°C	S/N
11		Cell lysate soluble
12		Cell lysate Insoluble
13	Post-induction 20hrs at 16°C	S/N
14		Cell lysate soluble
15		Cell lysate Insoluble
16	Post-induction 3hrs at 37°C	S/N
17		Cell lysate soluble
18		Cell lysate Insoluble
19	Post-induction 20hrs at 37°C	S/N
20		Cell lysate soluble
21		Cell lysate Insoluble

Table 6.13, Protein expression trials for the production soluble ECE-1 in E. coli. Conditions varied in temperature and induction time and samples were taken of the supernatant, and soluble and insoluble cell lysate fractions. These samples were then analysed using SDS-PAGE and western blot.

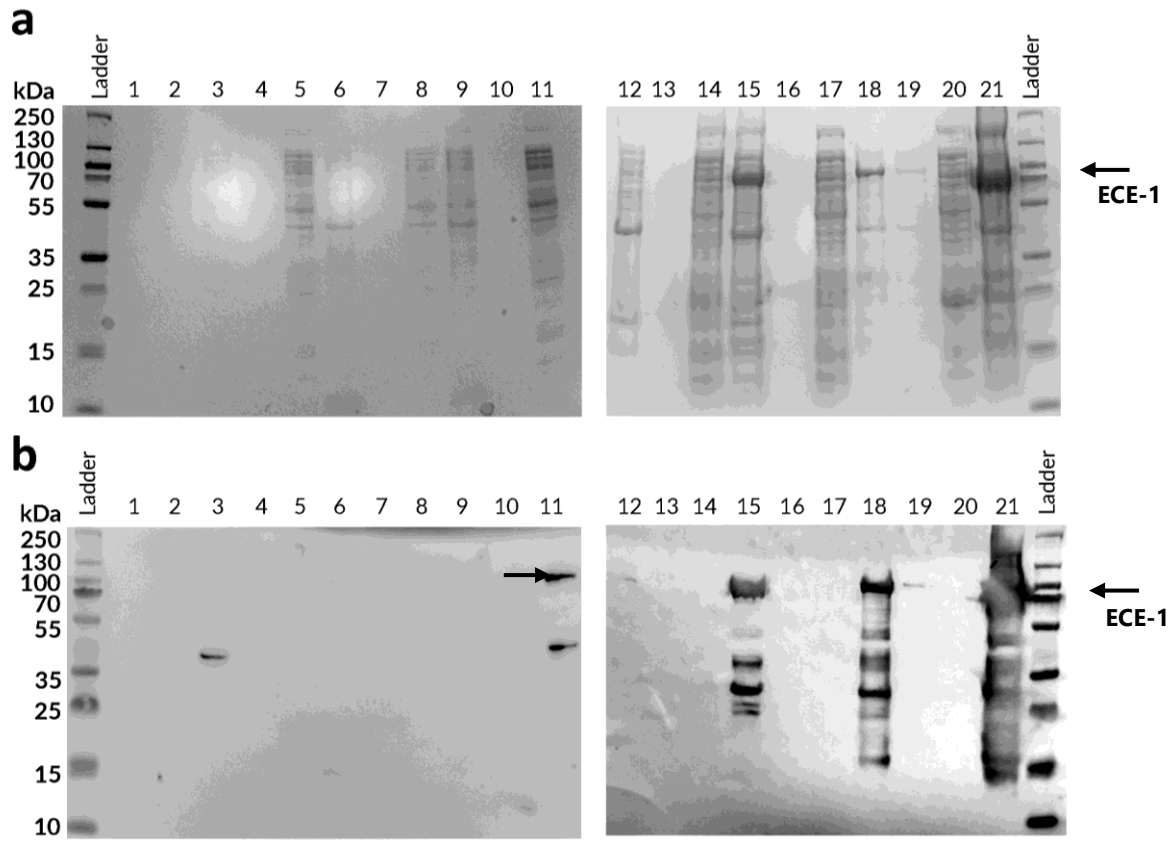


Figure 6.12, Soluble ECE-1 expression trials in E. coli RGB cells. (a) SDS-PAGE analysis of ECE-1 supernatant and soluble and insoluble cell lysate samples. (b) western blot analysis of ECE-1 supernatant and soluble and insoluble cell lysate samples. Sample identities are shown in Table 6.13 and bands corresponding the ECE-1 are marked with arrows.

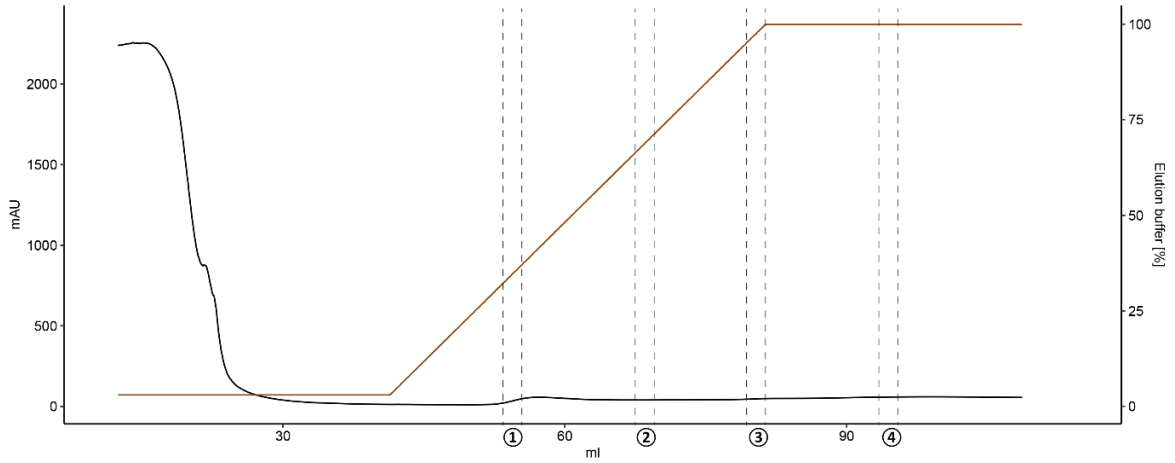


Figure 6.13, Histidine affinity trap purification of soluble ECE-1. ECE-1 was expressed in RGB cells using a 3-hour induction at 16°C. Resuspended cell lysate from 750 mL LB was loaded onto a histidine affinity trap column (GE

healthcare). Protein elution was achieved using a gradient of elution buffer over 40 mL. Over the course of the gradient elution there was an absence of clear peaks in the UV trace. Four samples were taken from fractions with slightly elevated UV absorbances. These samples were then analysed with SDS-PAGE and western blot.

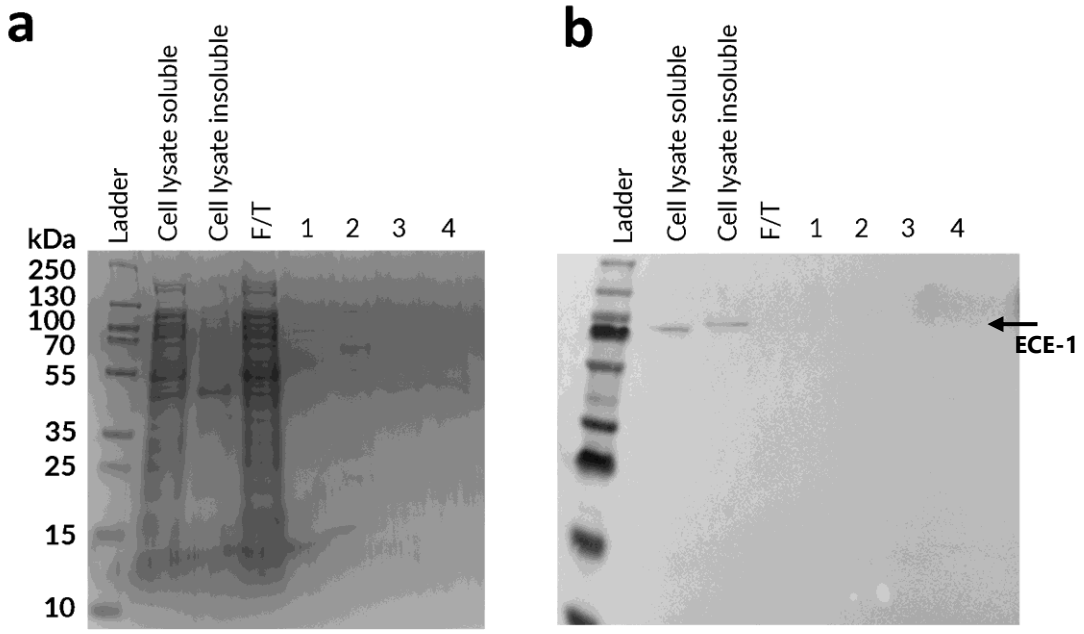


Figure 6.14, Soluble ECE-1 expression and purification in E. coli RGB cells. (a) SDS-PAGE and (b) western blot analysis of soluble and insoluble cell ECE-1 cell lysate post-induction and histidine affinity trap fractions from Figure 6.13. Bands corresponding to ECE-1 are marked with an arrow.

While soluble protein was present from western blot analysis, the quantity of protein was too low for purification and assessment of peptide cleavage. It was also unlikely that optimisation or scaling up of the protocol would have allowed the production of quantities required for structural work. These results, in combination with the inclusion body protein expression described in Schulz *et al.* (2009), strongly suggests a bacterial expression system is inappropriate for soluble ECE-1 expression.

6.4.3 ECE-1 expression in P. pastoris

For ECE-1 expression in *P. pastoris* it was necessary to remove the ECE-1 C428S gene from pET22b and ligate it into the yeast specific vector pPICzαA vector (Figure 6.15). Two PCR reactions were used to amplify ECE-1 C428S from the pET22b adding a 5′ 6x histidine tag, 5′ EcoRI site and a 3′ NotI site. The PCR product was then ligated into the vector pPICzαA (Figure 6.16). This specific vector allowed integration of a target gene into the *P. pastoris*

genome for secreted protein expression. After integration of pPICzαA ECE-1 C428S into *P. pastoris,* clones where screened for protein expression.

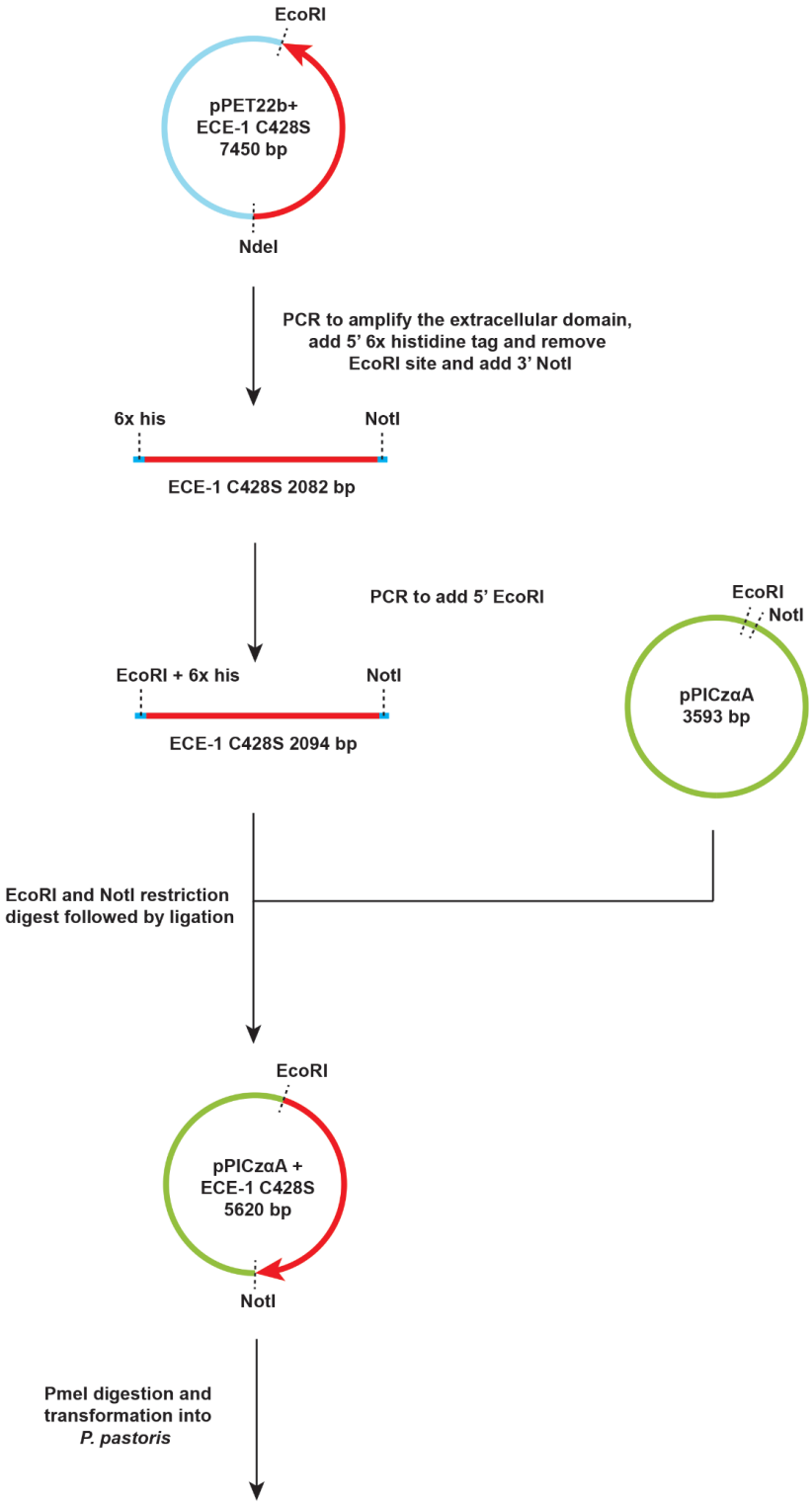


Figure 6.15, Cloning plan to generate ECE-1 C428S pPICzαA. The ECE-1 C428S gene was extracted from pET22b and ligate it into pPICzαA for protein expression in P. pastoris.

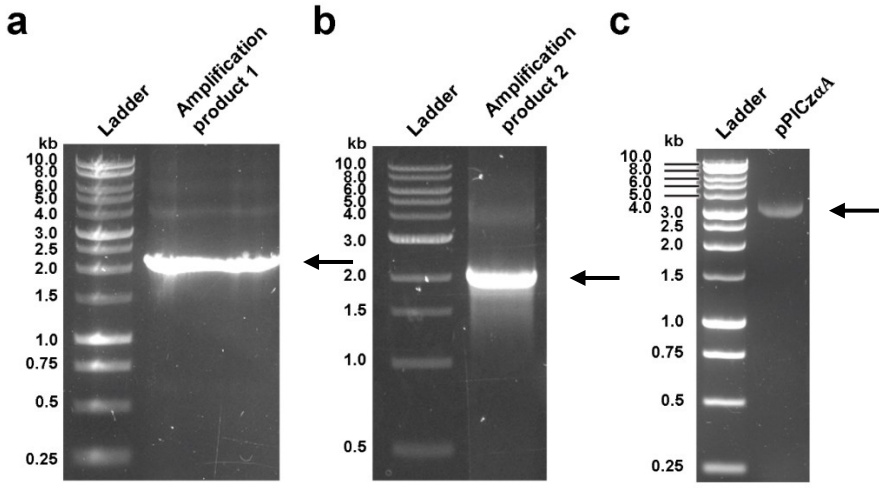


Figure 6.16, Production of ECE-1 C428S pPICzαA. (a) DNA product from a PCR amplifying ECE-1 C428S from pET22b and adding a 5' 6x histidine tag. (b) DNA product from a second PCR adding a 5'EcoRI site for ligation into pPICzαA. (c) pPICzαA vector.

19 *P. pastoris* clones were selected for small-scale protein expression experiments. As protein was expressed with a secretion tag only post-induction supernatant samples were analysed for ECE-1 expression. Samples were analysed using western blot to determine if ECE-1 was expressed and which clones had the highest expression levels (Figure 6.17). Translated ECE-1 protein has a mass of ~80 kDa. However due to previous observations of mass increase in NEP (Chapter 3), it was expected that ECE-1 could exhibit similar behaviour due to glycosylation. Several clones contained a visible protein band on the western blot at ~100 kDa corresponding to ECE-1. Two clones, 7 and 11, had far higher levels of protein expression in supernatant samples and were selected for large-scale ECE-1 expression.

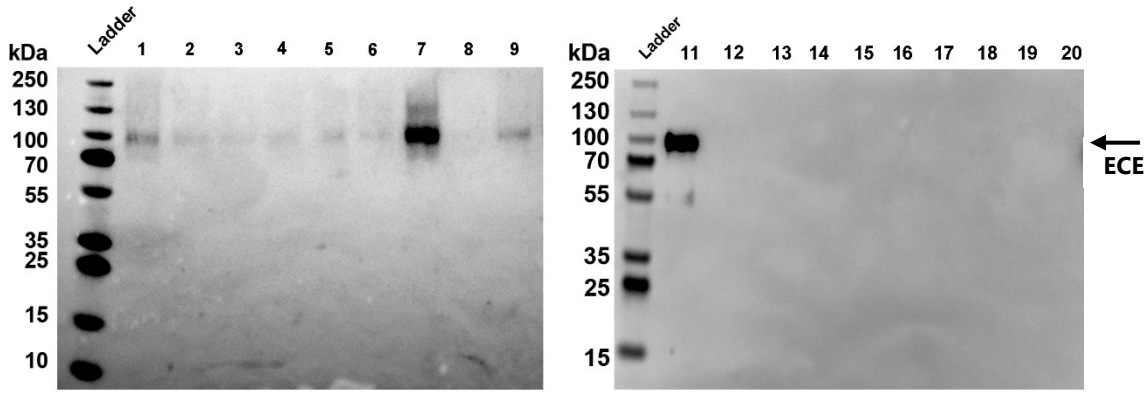


Figure 6.17, small-scale ECE-1 C428S P. pastoris expression cultures. 19 post-induction supernatant samples from P. pastoris clones were screened using western blot analysis to determine which clones had the highest level of ECE-1 C428S protein expression. Bands corresponding to ECE-1 expression are marked with arrows.

The protocol for large-scale ECE-1 expression in *P. pastoris* was initially performed in an identical way to that of NEP expression. However, western blot analysis of post-induction samples revealed a lower induction time of 24 hours was required as longer induction times resulted in substantial protein degradation (Figure 6.18).

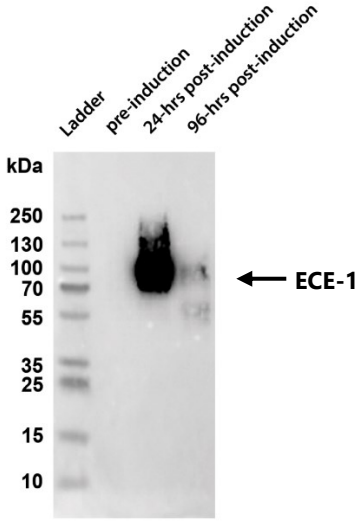


Figure 6.18, Large-scale ECE-1 C428S expression in P. pastoris clone. Supernatant samples were taken pre-induction and post-induction at 24 and 96 hours and analysed using a western blot. A band corresponding to ECE-1 was present in the 24 hours post-induction sample and is marked with an arrow.

The protocol for large-scale ECE-1 expression was then repeated but with a reduced 24-hour induction period. Post-induction supernatant harvested and loaded onto a histidine trap affinity column with a 10 mM imidazole wash step followed by a 250 mM imidazole

elution step. A single elution peak was observed (Figure 6.19) and peak containing fractions were pooled and concentrated before loading onto a superdex size exclusion column. A protein peak eluted from the size exclusion column at 75-85ml. The same peak was observed during NEP purification, however the peak corresponding to ECE-1 was substantially lower in absorbance (Figure 6.20). Samples were taken throughout the expression and purification and were analysed with SDS-PAGE and western blot (Figure 6.21).

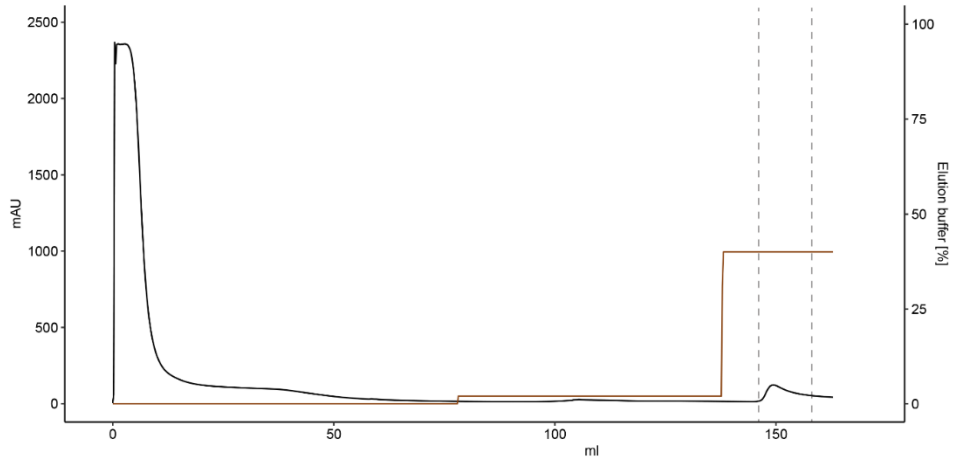


Figure 6.19, P. pastoris expressed ECE-1 C428S histidine affinity trap purification. 500 mL of protein containing supernatant from a P. pastoris culture was loaded onto a histidine affinity trap column (GE healthcare). Once loaded the column was washed with binding buffer followed by a wash step of 2% elution buffer. Protein elution was achieved at 50% elution buffer. A single peak was observed and peak containing fractions (indicated between dash lines) were pooled.

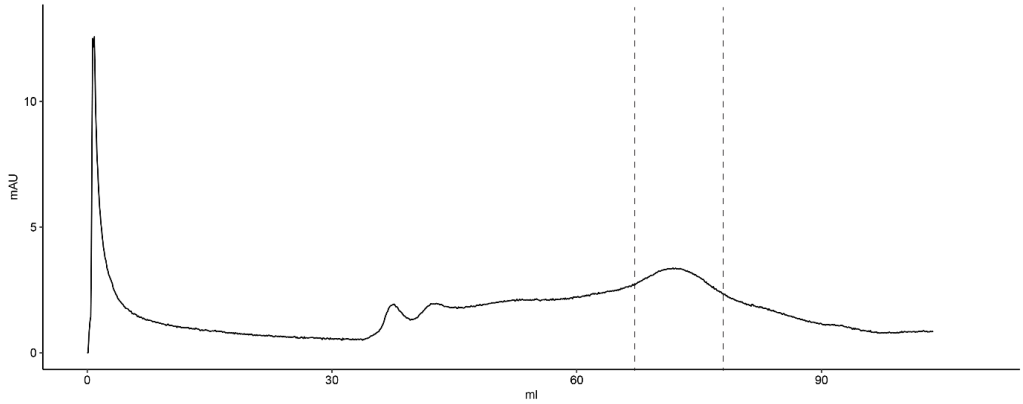


Figure 6.20, P. pastoris expressed ECE-1 C428S size exclusion purification. The pooled fractions eluted from the histidine affinity trap column were loaded onto a Superdex 16/60 size exclusion column to remove impurities from

the induction media and to buffer exchange the protein into binding buffer. The peak containing fractions (indicated between dash lines) were pooled.

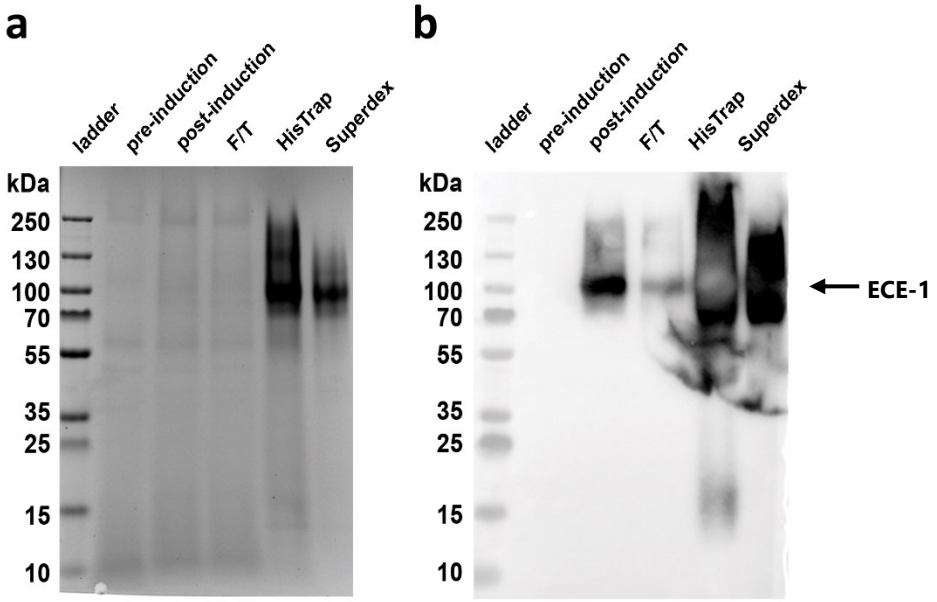


Figure 6.21, P. pastoris ECE-1 C428S expression and purification. (a) SDS-PAGE and (b) western blot analysis of supernatant pre- and post-induction samples taken from the P. pastoris expression culture just before induction and 24-hours post induction. Samples of the histidine affinity column flow through (F/T) are also shown. The eluted and pooled fractions after purification steps reveal the high purity of the final product (mark with arrows). Smearing of the final sample can be attributed to protein glycosylation.

Protein expression in the supernatant of *P. pastoris* was clearly observed from the western blot. While a small level of protein did not bind to the histidine column and was observed in the flow through, the majority of protein appear to bind to the column. The histidine trap did provide purification of ECE-1 and it was the only major protein band present in the eluted sample (Figure 6.21). The protein band corresponding the ECE-1 displayed a high level of smearing both in the SDS-PAGE and western blots. The most likely cause of which was glycosylation of the protein. After the second purification step the sample was observed to contain the same ~100 kDa protein band but with reduced smearing. This is likely a consequence of removing some high glycosylation species. The absorbance of the final purified product was measured to determine an approximate yield of 0.2 mg/L.

This purified sample was then used to for a cleavage assay. Full kinetic characterisation for the ECE-1 sample was not performed. While full characterisation would have been preferential, assessment of a single concentration was used to provide insights into whether *P. pastoris* was an appropriate expression system. Assays were conducted at a single substrate concentration and the initial rate for that concentration is shown in Figure 6.22.

Substrate cleavage was confirmed for ECE-1 expressed in *P. pastoris*. The initial rate for the single ECE-1 concentration was observed to be similar to that of the NEP control. However, it should be noted that the protein sample of ECE-1 and NEP were assayed at 83.2 nM and 6.25 nM respectively. While initial rate was lower than that observed for NEP, taking into account the protein concentration, the rate was substantially higher than that previously determined for ECE-1 expressed in *E. coli* RGB cells and refolded.

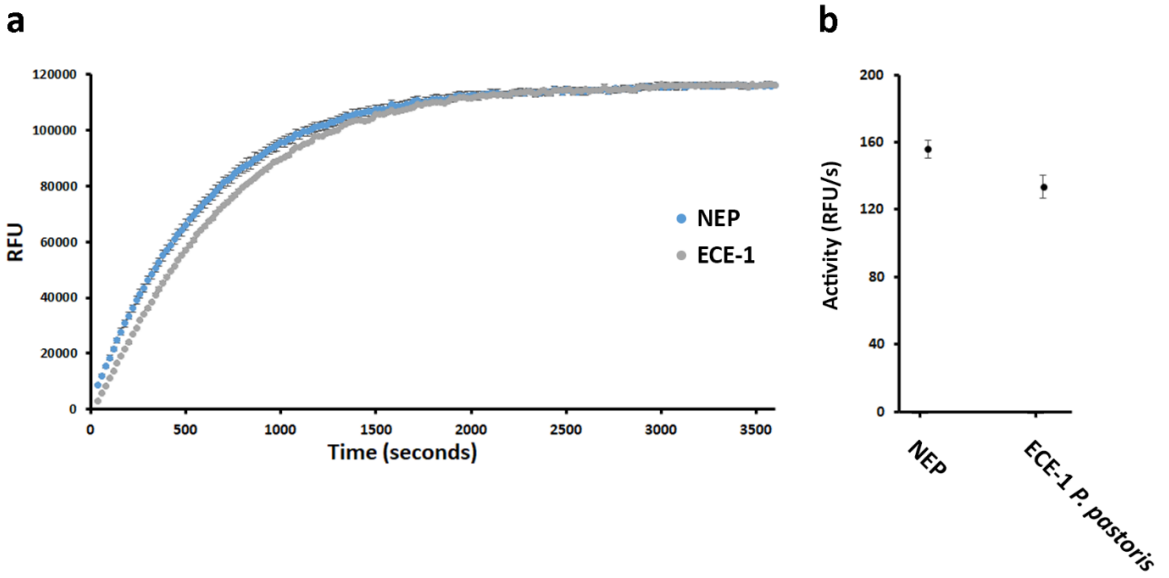


Figure 6.22, Peptide cleavage assay for ECE-1 expressed in P. pastoris. (a) Peptide cleavage was assessed at a single substrate concentration (6.25 nM for NEP and 83.2 nM for ECE-1). Substrate Mca-RPPGFSAFK-(Dnp) was used at 10 µM final concentration. Substrate cleavage by NEP and ECE-1 was repeated three times with the mean and ±SD error bars shown above. (b) Initial rates of cleavage for protein samples at the single concentration. The mean and ±SD error bars are shown for the initial rates calculate from three repeats.

6.4.4 ECE-1 expression in human embryonic kidney cells

For ECE-1 expression in HEK 293T cells it was necessary to remove the ECE-1 C428S gene from pPICzαA and ligate it into the mammalian vector pOPINTTGneo vector (Figure 6.23). A PCR reaction was used to amplify ECE-1 C428S from the pPICzαA adding a 5′ KpnI site and a 3′ PmeI site. The PCR product was then ligated into the vector pOPINTTGneo (Figure 6.24). This specific vector allowed transient and secreted protein expression in mammalian cells.

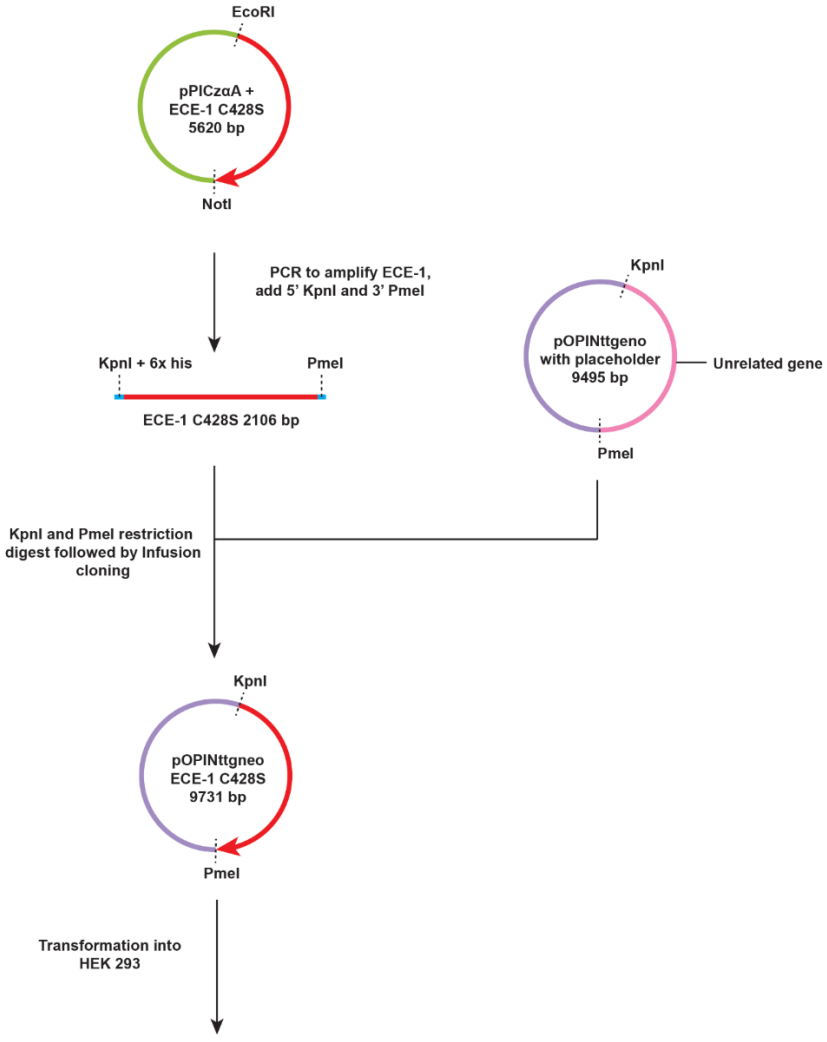


Figure 6.23, Cloning plan to generate ECE-1 C428S pOPINTTGneo. The ECE-1 C428S gene was extracted from pPICzαA and ligate it into pOPINTTGneo for protein expression in HEK 293T cells.

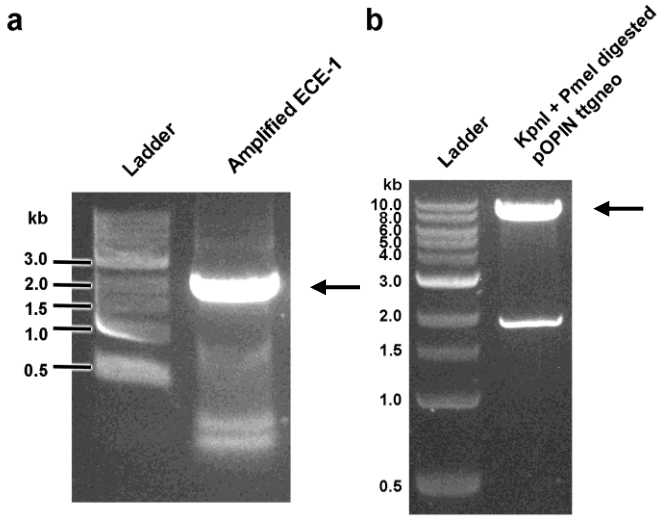


Figure 6.24, Production of ECE-1 C428S pOPINTTGneo. Agarose gel electrophoresis analysis of the (a) DNA product from a PCR amplifying ECE-1 C428S from pPICzαA and (b) pOPINTTGneo vector double digested to remove an unrelated placeholder gene.

The final pOPINTTGneo ECE-1 C428S construct was transformed into HEK 293T cells and a small-scale expression trial was used to determine the best transformation protocol for protein expression through variation in DNA and polyethylenimine (PEI) concentrations (Table 6.14). 24-hours after transformation the supernatant from small-scale expression trials were analysed with SDS-PAGE and western blot (Figure 6.25). As previously stated, the translated ECE-1 protein is ~80 kDa but with glycosylation this figure is expected to be larger. ECE-1 expressed in *P. pastoris* was observed to be ~100 kDa due to glycosylation (Figure 6.21). In HEK 293T cells the level of glycosylation was expected to be of a similar level (Croset *et al.,* 2012). All supernatant samples displayed a band at ~130 kDa on the western blot corresponding to ECE-1. Of transduction conditions the greatest ECE-1 band intensity was observed in conditions 1 and 3.

Condition	DNA	PEI	DNA:PEI
1	4µg	4µg	1:1
2	2µg	4µg	1:2
3	4µg	8µg	1:2
4	2µg	6µg	1:3
5	4µg	12µg	1:3

Table 6.14, Transfection ratios for protein expression trials of ECE-1 C428S in HEK 293T cells. DNA and PEI concentrations and ratios were varied for 500 µL cultures to identify a condition for optimal protein expression.

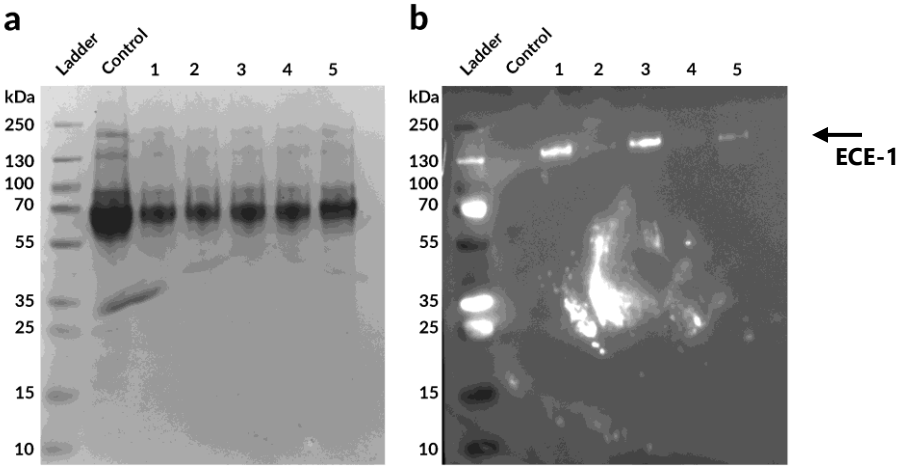


Figure 6.25, Small-scale trials of ECE-1 C428S expression in HEK 293T cells. (a) SDS-PAGE and (b) western blot analysis of cell supernatant 24-hours after induction. An untransformed control is shown with the five different transformation conditions listed in Table 6.14. Bands corresponding to ECE-1 were present in the western blot and are marked with an arrow.

Large-scale expressions were performed using the scaled up the transformation condition 1 in a 2 m^2 roller bottle. Transformed cells were incubated for four days, after which the protein containing supernatant was extracted and protein purification was conducted. Protein purification was performed first with a histidine strap affinity column. Supernatant was filtered and loaded onto the histidine trap column and eluted using a gradient of imidazole containing buffer (Figure 6.26).

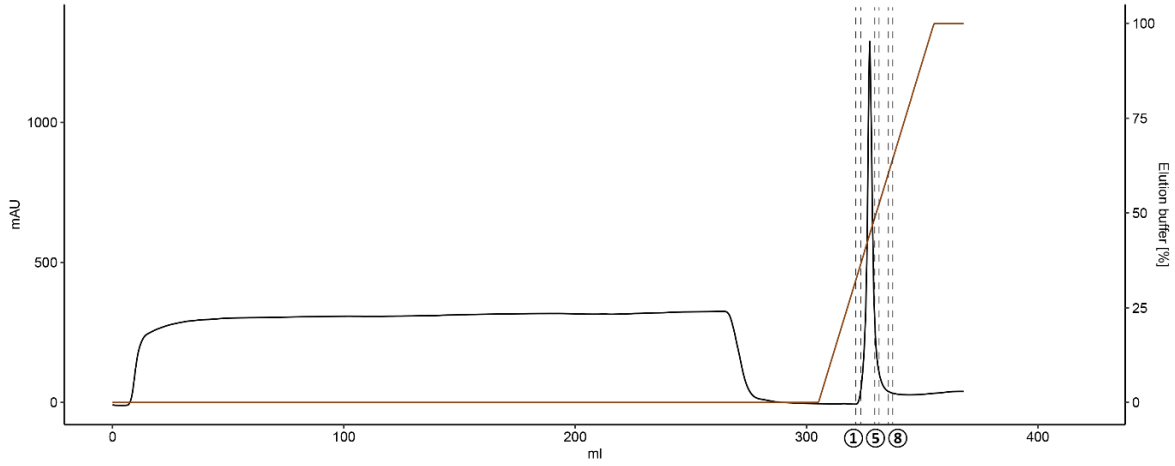


Figure 6.26, HEK 293T large-scale ECE-1 histidine affinity trap purification. 500 mL of protein containing cell supernatant from a HEK 293T culture was loaded onto a histidine affinity trap column (GE healthcare). Protein elution was achieved using a gradient of elution buffer over 40 mL. Over the course of the gradient elution a single

sharp peak was observed. Eight fractions were collected across the peak and were analysed with SDS-PAGE and western blot.

A single peak was observed during elution and fractions from that peak were analysed by SDS-PAGE and western blot (Figure 6.27). It was clear that only fractions corresponding to the second half of the peak contained ECE-1 protein as indicated from the western blot. The presence of other protein bands in addition to the band corresponding to ECE-1 indicated that further purification was required.

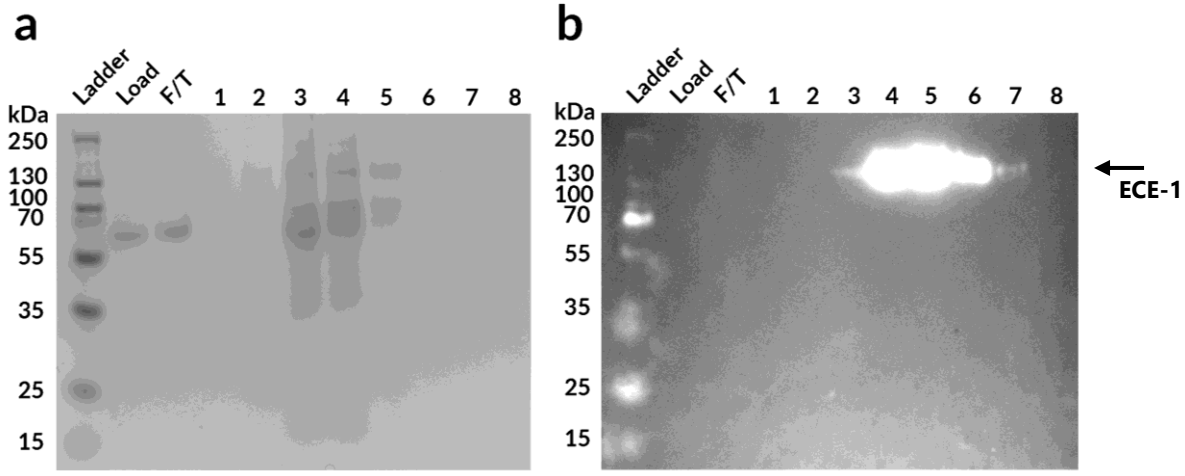


Figure 6.27, Protein containing fractions from a histidine affinity purification of ECE-1 C428S expression in HEK 293T cells. (a) SDS-PAGE and (b) western blot analysis loaded supernatant, flow through (F/T) and eight fractions collected within the eluted peak. Fractions 3-7 contained a band at ~130 kDa corresponding to ECE-1 which is mark with an arrow.

Protein containing fractions 3-7 were pooled and loaded onto a superdex size exclusion column for further purification. This had the additional benefit of removing any imidazole in the protein sample. Four peaks were present in the size exclusion elution profile (Figure 6.28). Fractions from these peaks were pooled and analysed using SDS-PAGE and western blot to reveal that peak three contained ECE-1 (Figure 6.29). Given the overlap between the elution of peak 3 and peaks 2 and 4 the final pooled peak 3 protein sample contained several weaker bands in addition to ECE-1. This made estimation of the quantity of purified protein difficult but was estimated at ~0.2 mg/L.

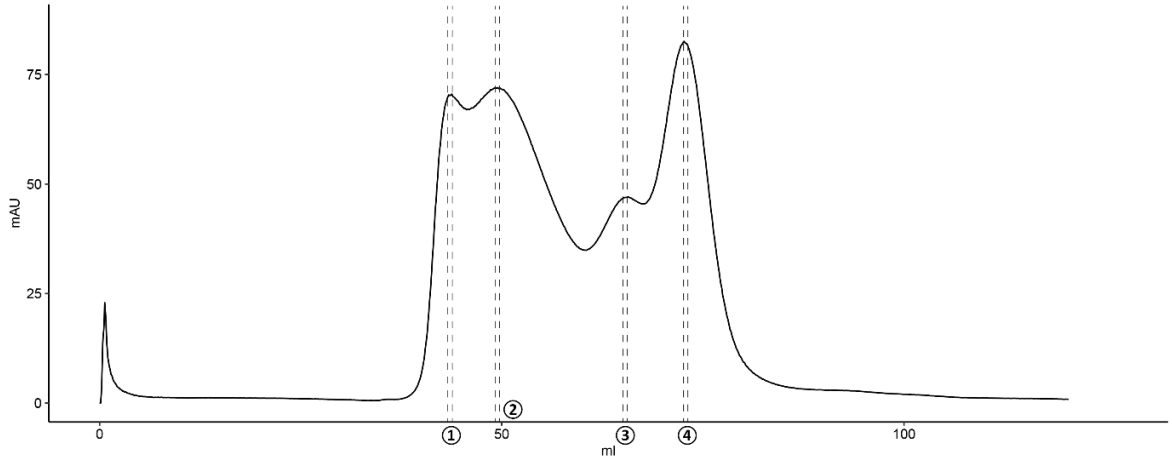


Figure 6.28, Size exclusion purification of ECE-1 C428S expressed in HEK 293T cells. The pooled fractions (3-7) eluted from the histidine affinity trap column were loaded onto a Superdex 16/60 size exclusion column to remove impurities from the induction media and to buffer exchange the protein into binding buffer. Four peaks were observed and peak containing fractions (indicated between dash lines) were pooled and analysed with SDS-PAGE and western blot.

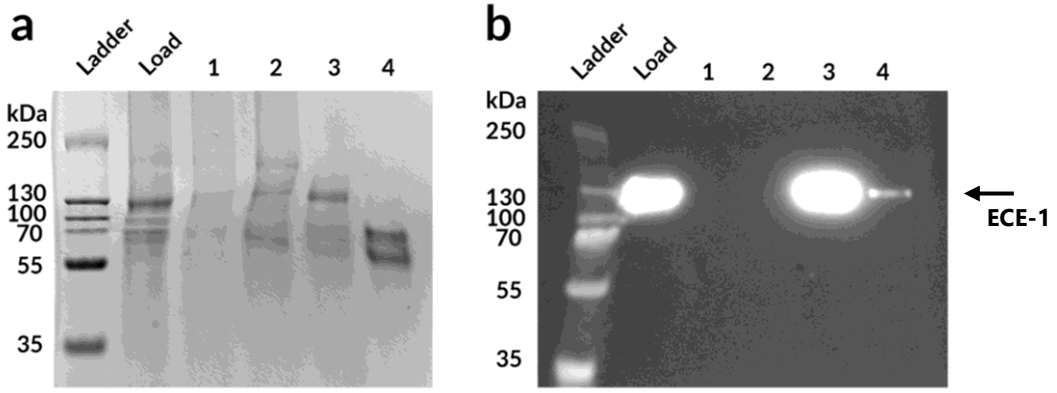


Figure 6.29, Protein containing fractions from a size exclusion purification of ECE-1 C428S expression in HEK 293T cells. (a) SDS-PAGE and (b) western blot analysis loaded protein sample for the histidine affinity purification and pooled fractions from the four eluted peaks. Fraction 3 contained a band at ~130 kDa corresponding to ECE-1 which is marked with arrows.

The pooled protein from peak 3 was then used to for a substrate cleavage assay. Assays were conducted at a single substrate concentration and the initial rate for that concentration is shown in Figure 6.30. Substrate cleavage was confirmed for ECE-1 expressed in HEK 293T cells. The initial rate for the single ECE-1 concentration was observed to be similar to that of the NEP control. However, the initial rate determined for the single ECE-1 concentration was far lower than that of the NEP. It should also be noted that the protein sample of ECE-1 and NEP were assayed at 833 nM and 6.25 nM respectively.

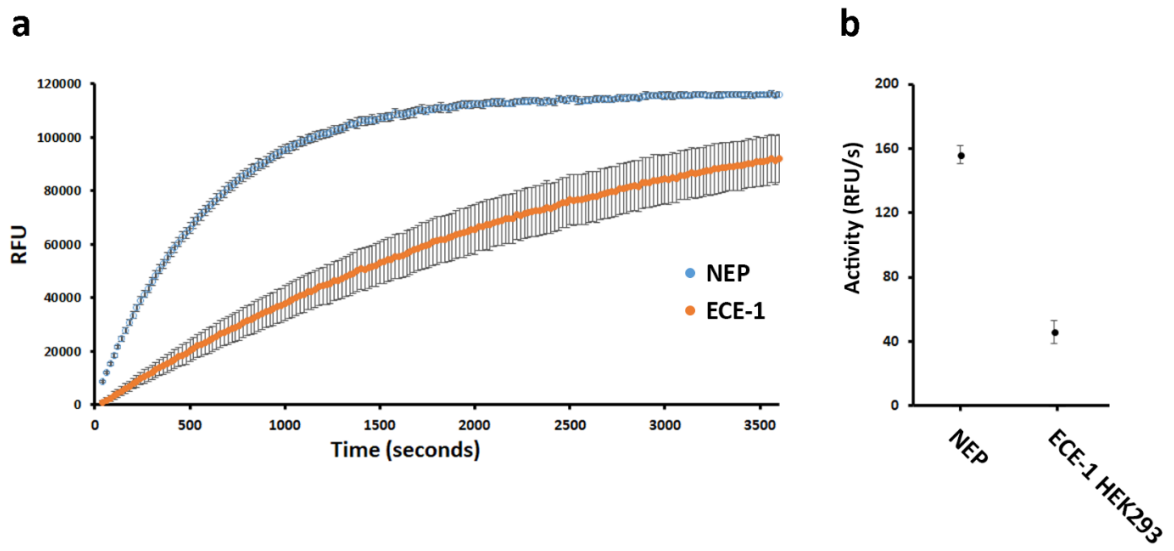


Figure 6.30, Peptide cleavage assay for ECE-1 expressed in HEK 293T cells. (a) Peptide cleavage was assessed at a single substrate concentration (6.25 nM for NEP and 833 nM for ECE-1). Substrate Mca-RPPGFSAFK-(Dnp) was used at 10 µM final concentration. Substrate cleavage by NEP and ECE-1 was repeated three times with the mean and ±SD error bars shown above. (b) Initial rates of cleavage for protein samples at the single concentration. The mean and ±SD error bars are shown for the initial rates calculate from three repeats.

6.5 Conclusions

The initial aim for the work on ECE-1 was to establish an expression system in which active protein could be expressed and used for structural characterisation. Schulz *et al*. (2009) demonstrated that ECE-1 could be generated in a bacterial expression and refolded for use in structural studies. This method was investigated at the start of this chapter with the aim of replicating ECE-1 production and providing evidence that the protocol was suitable for production of active protein despite the absence of glycosylation.

The method of ECE-1 expression described by Schulz *et al*. (2009) did not include information regarding the refolding process, instead a refolding screen was used to determine a suitable condition. While some refolding conditions produced soluble protein that was able to cleavage peptide substrate, the observed cleavage was extremely low. Additionally, the purification of refolded samples indicated that refolding had generated a heterogenous array of differently folded species which were not suitable for structural work.

Indeed, attempts were made to express soluble ECE-1 protein in *E. coli* RGB cells, but it was clear only small quantities of protein were produced that could not be purified.

Additional expression systems were attempted with the aim of generating soluble ECE-1. A clear place to start, based on the success of NEP, was the yeast expression system *P. pastoris*. This expression system offered more sophisticated post-translational processing including protein glycosylation. 19 transformed clones of *P. pastoris* were screen for ECE-1 expression and the clone with the highest level of ECE-1 expression was selected for large-scale protein production. Expressed ECE-1 was purified by histidine trap affinity chromatography with a further size exclusion chromatography step to yield pure protein. This pure protein was then analysed using the fluorescent assay to measure initial rate data.

While initial rate was only determined for a single ECE-1 concentration it was used to provide a measure of the efficiency of the expression system in producing active protein. It also provided a means of comparing the different protein expression systems tested. The observed initial rate for ECE-1 expressed *P. pastoris* was lower than that of the NEP control but was far higher than that determined for the ECE-1 express in *E. coli* and then refolded.

In addition to *P. pastoris,* a mammalian expression system in the form of human embryonic kidney 293T cells was performed for ECE-1 expression. Screening transformation protocols for HEK 293T cells allowed selection of an optimal protocol for protein expression. ECE-1 expressed in HEK 293T cells could be purified by histidine trap affinity chromatography with a further size exclusion chromatography step. The final product analysed by SDS-PAGE and western blot contained a band for ECE-1 but also several other bands corresponding to protein impurities. Despite this, the final protein sample was analysed using the fluorescent assay to measure initial cleavage rate data. Again, a single protein concentration of the ECE-1 sample was assayed, and peptide cleavage was detected. The initial rate data was lower than that of the NEP control in additional to ECE-1 produced in *P. pastoris*. Initial rate values for expressed ECE-1 are summarised in Table 6.15.

	Protein concentration (nM)	Initial rate (RFU/s)	Standard deviation (RFU/s)

ECE-1 *P. pastoris*	83.20	133.63	1.58
ECE-1 HEK 293T	833.00	45.91	7.00
NEP control	6.25	155.95	5.56

Table 6.15, ECE-1 peptide cleavage assay initial rate data. Table summarises the initial rate data from Figures 6.22 and 6.30.

While characterisation of peptide cleavage for ECE-1 was simplistic, and a direct comparison cannot be made due to the variation in protein concentration, the calculated initial rate data does suggest *P. pastoris* may be a better system for production of active ECE-1 (Table 6.15). It should however be noted that the presence of impurities in the purified ECE-1 sample from HEK293T cells may affect the initial rate data.

The expression trials discussed above have revealed that soluble ECE-1 could be expressed in both *P. pastoris* and HEK 293T expression systems and that protein produced possessed protease activity as demonstrated in the peptide cleavage assay. While protein could be expressed, there was a major limitation in that both expression systems generated a low level of protein. Yields of ~0.2 mg/L for both expression systems were of an insufficient quantity to begin structural characterisation.

Time limitations in the project prevented further work on ECE-1, however, from the work conducted *P. pastoris* emerged as the most promising expression system for several reasons. First, the protein expressed in *P. pastoris* could be purified with relative ease to yield a highly pure final sample. Second, the highest initial rate peptide cleavage was recorded for ECE-1 expressed in this system. Third, while expression levels of ~0.2 mg/L were insufficient to begin structural characterisation, with greater time it would be possible to revisit the *P. pastoris* transformation and screen more clones with the aim of identifying a clone with greater expression levels.

While time restrictions prevented structural characterisation of ECE-1, information regarding substrate binding in ECE-1 can be gained from the previously reported ECE-1 phosphoramidon-bound structure. Below a small section is included to comment on the current structural understanding of ECE-1 specificity. This brief section provides the opportunity to present some new insights into ECE-1 using the published structure of the

ECE-1 extracellular domain in conjunction with NEP structures discussed in Chapters 3 and 5.

6.5.1 ECE-1 structural analysis

The extracellular domains of ECE-1 and NEP are highly similar with an RMSD of 1.4 Å (Figure 6.31). Indeed, one of the most highly conserved regions of the protein is the active site (residues that form the S1-S2′ sites). A comparison between the active site of phosphoramidon-bound ECE-1 (3DWB) and NEP (1DMT) reveals the almost identical position of both the inhibitor and the surrounding residues (Figure 6.32). Because of this almost identical binding interaction, it should be unsurprising that ECE-1 and NEP have similar catalytic efficiencies for many substrates (Johnson and Ahn, 2000) and share a preference for cleavage at the amino side of a hydrophobic residue located in the S1′ site (Rawlings *et al.,* 2018).

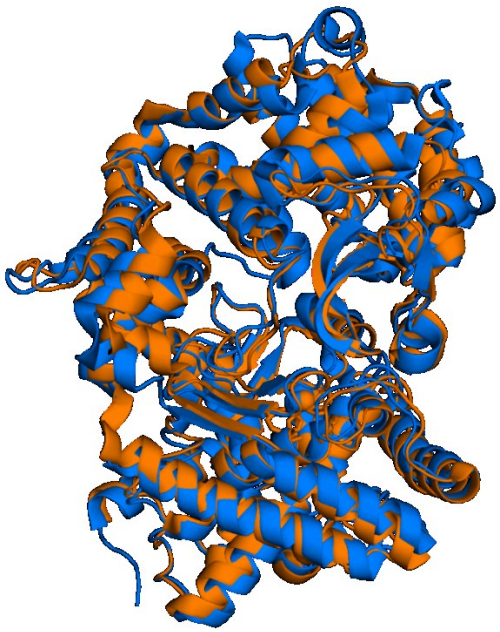


Figure 6.31, Ribbon representation of ECE-1 and NEP superposition. The extracellular domain of ECE-1 (orange) and NEP (blue) display a high degree of structural similarity with an RMSD of 1.424 (calculated using the Align function in Pymol).

Given the high level of conservation in active site residues and the similarity in phosphoramidon binding (Figure 6.32) it is likely that many of the substrate binding interactions discussed with regards to the peptide-bound NEP structure in Chapter 5 also

hold true for ECE-1. That is, the major drivers of substrate backbone interactions in the S1-S2′ sites for ECE-1 are likely contributed by residues Asp566, Ala567 and Arg738.

The interactions within the active site for ECE-1 and NEP are almost identical and account for the similarities seem in preferred substrate cleavage position. It seems unlikely, however, given this similarity, that it is the active site residues alone that account for the distinct specificities between ECE-1 and NEP when considering substrate identity. This discrepancy is most drastic when considering small (<7 residue) peptides for which ECE-1, unlike NEP, has very low affinity for (Johnson *et al.,* 1999).

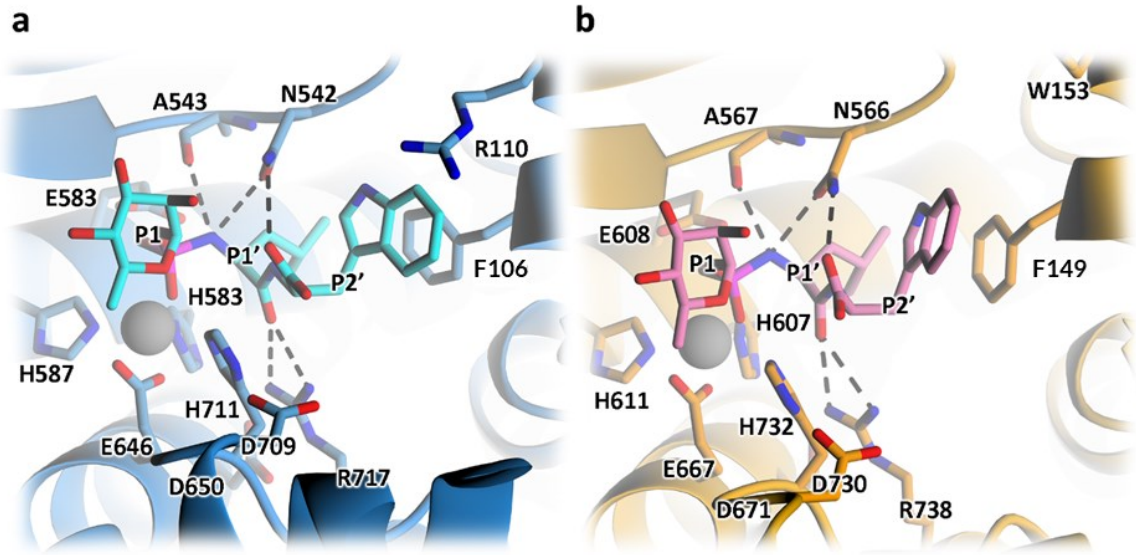


Figure 6.32, Active site binding of inhibitor phosphoramidon in NEP and ECE-1. Active site residues in (a) NEP and (b) ECE-1 are shown with bound phosphoramidon (shown in cyan and pink respectively) with zinc ion shown in grey.

Substrate specificity can be considered to be the culmination of steric restraints and electrostatic/hydrophobic interactions (Náray-Szabó, 1993). Given the high degree of similarity in cavity size and active site spatial organisation for ECE-1 and NEP, it is unlikely that steric restraints can account for the distinct specificity between the two proteins. It is therefore likely that electrostatic or hydrophobic interactions plays a critical role.

Analysis of the surface electrostatic potential of ECE-1 and NEP internal cavities has revealed their distinct nature (Figure 6.33). Such electrostatic organisation has been suggested to be a major contributor to long-range positioning of substrates in other proteins (Zhang *et al.,* 2011). It is possible that the varied substrate specificity is rooted in the varied long-range

attraction of substrates due to the differences in the observed electrostatic potentials. It is also likely that the varied identity of residues lining the cavity in ECE-1 and NEP, which contribute to the observed different electrostatic potentials, interact with substrates upon binding. It may be these interactions, proximal to the active site, which account for the distinct substrate specificity.

The concept of proximal interactions has been discussed in Chapter 5 with the possible identification of an exosite in NEP. Proximal interactions in ECE-1 would offer an explanation for ECE-1's low cleavage efficiency of small peptides. Small peptides may be of insufficient length to interact with key residues in the ECE-1 cavity. It is clear that further structural characterisation of ECE-1 is required, however, the points discussed above suggests residues proximal to the active site in ECE-1 and NEP have importance in protein-peptide binding interactions.

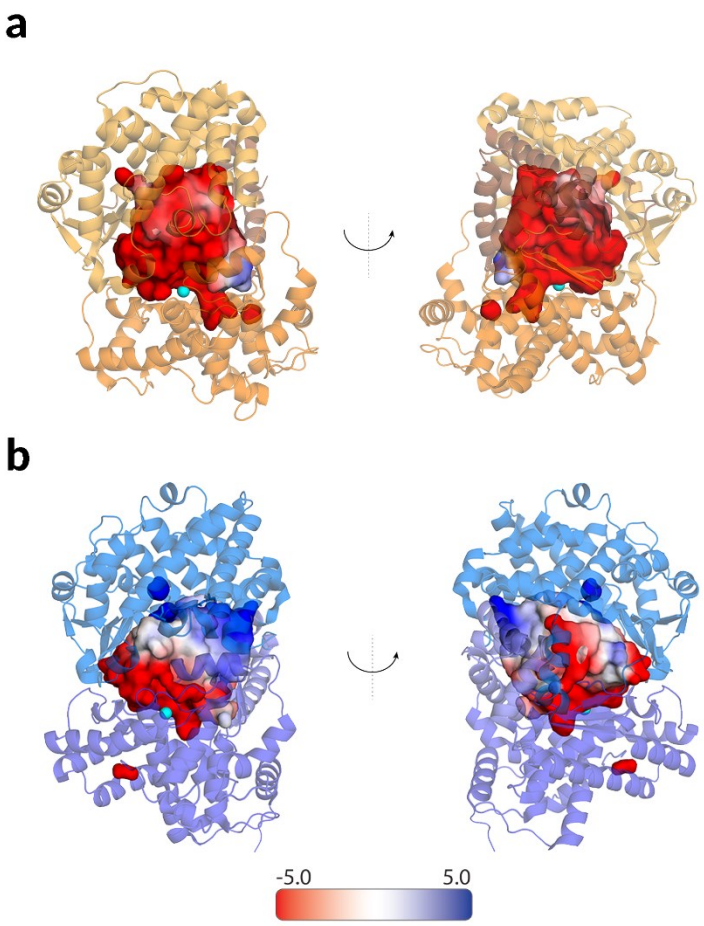


Figure 6.33, Ribbon representation of ECE-1 and NEP with electrostatic potentials of internal cavity surfaces. Both (a) ECE-1 and (b) NEP contain large central cavities with distinct electrostatic potentials (calculated with APBS). The active site is located within each respective cavity based around a zinc ion (cyan).

Chapter 7 – Discussion

7.1 Summary

Alzheimer's disease presents an immense hurdle for science in the 21st century. Whilst the last two decades have seen substantial progress made in the field of Alzheimer's research, it has further revealed the complex nature of the disease. It is clear that a greater understanding of the underlying processes of Alzheimer's disease are required before therapies can be developed that significantly reduce or halt disease progression.

The current understanding of Alzheimer's disease firmly points towards the peptide amyloid-β as a major contributor to disease progression. Because of this, the regulation of amyloid-β is a critical area of Alzheimer's disease research. Amyloid-degrading enzymes (ADEs) are known to be key drivers of amyloid-β degradation and research into these enzymes builds upon the knowledge of the underlying mechanisms of Alzheimer's disease.

To date, over 30 proteins have been identified as possible ADEs. These proteins exhibit different potencies, localisations and many display additional roles. The proteins that were studied and described in this thesis, neprilysin (NEP) and endothelin converting enzyme I (ECE-1), are highly potent ADEs but also have an additional function in blood pressure regulation. The involvement of NEP and ECE-1 in amyloid-β degradation and blood pressure regulation makes them high-value therapeutic targets for Alzheimer's disease and cardiovascular disease.

The aim of this Ph.D. research was to provide structural and functional characterisation of NEP and ECE-1. Through this characterisation the mechanism of NEP and ECE-1 in peptide degradation could be further revealed and their potential as therapeutic targets could be realised. The discussion below is organised into sections based on the protein studied.

7.1.1 Neprilysin

While NEP is the best-known ADE (Iwata *et al*. 2001), the majority of structural and functional characterisation has been focused on its role in blood pressure regulation. Inhibition of NEP is desirable in patients with high-blood pressure, and a number of crystal structures have been reported for NEP with inhibitors bound. These structures have served as valuable starting points for functional characterisation, revealing the overall protein fold and the location of the active site within a large central cavity.

Despite the number of inhibitor-bound structures reported, at the start of this project a substrate-free form had yet to be determined. With any protein, determination of the native or substrate-free structure provides an insight into its unbound state. This unbound state is critical to identify conformational movements that may result from ligand binding. These conformational movements can be large domain movements or can be conformational changes in specific amino acid residues. It is often these movements that are key to the molecular mechanism underlying a protein's function.

Chapter 3 details the processes used to determine the first substrate-free NEP structure (Figure 7.1). This structure revealed the protein conformation and position of active site residues in the absence of peptide substrate. Previous inhibitor-bound structures of NEP have determined that the protein adopts a closed conformation where the central cavity is occluded and inaccessible to substrates. The substrate-free NEP structure was also revealed to adopt a similar closed conformation indicating that inhibitor binding was not required for the protein to adopt this closed form. Through comparisons to inhibitor-bound structures it was also possible to identify the highly similar positions occupied by active site residues. This observation suggests the active site is static in nature and does not display major conformational movements upon ligand binding.

The conformation of substrate-free NEP raised question regarding how substrates access the active site and whether the observed closed conformation was an artefact of the crystal environment or a true picture of its natural state. To address this important question in NEP and to validate the closed structure observed in the crystal structure, small-angle X-ray scattering technique was used. Experimentally calculated scattering parameters R_g and D_{max} were compared to predicted scattering parameters from the NEP crystal structure. These scattering parameters were highly similar and indicated that in solution NEP adopts a similar conformation to the one observed in the crystal structure. Indeed, attempts were made to elicit conformational flexibility of subdomains using varation in buffer conditions, however, no significant movement was observed.

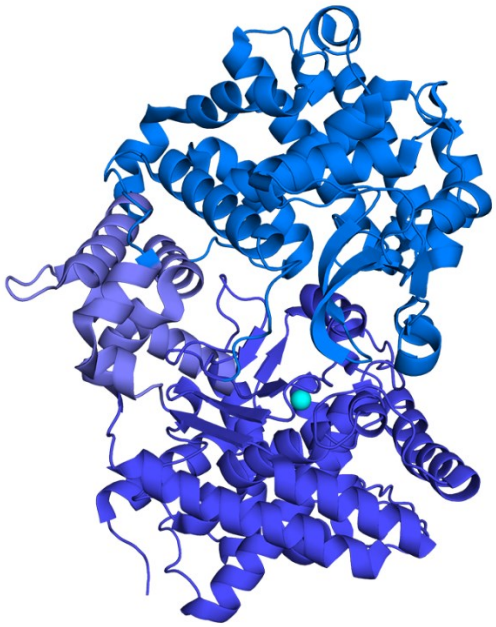


Figure 7.1, Crystal structure of the extracellular domain of NEP in a substrate-free form (6GID). The protein contains a large central cavity where the active site is located based around a zinc ion (cyan).

Further crystallographic research presented in Chapter 5 detailed the methods used to determine the first peptide-bound structure of NEP. In order to obtain such a structure, removal of catalytic activity was required to prevent NEP degrading peptides during the incubation period required for crystallisation.

Catalytic inactivation for crystallisation had been successfully used in other zinc metalloproteases including insulin-degrading enzyme (Manolopoulou *et al.,* 2007). Our results with NEP have revealed that mutation of the conserved glutamate in the active site motif HExxH removed detectable protease activity. Similar results have been reported for the model zinc metalloprotease thermolysin (Vazeux *et al.,* 1996) for which detailed characterisation of the catalytic mechanism has been achieved. The conserved active site motif and similar response to glutamate mutation provides further evidence that NEP utilises a catalytic mechanism analogous to thermolysin.

Using circular dichroism and X-ray crystallography it was possible to confirm that mutant NEP was near-identical in structure to wild-type NEP and was suitable model to study peptide-binding interactions. This protein was then used to determine the first peptide-bound structure of NEP in complex with vasoactive peptide CNP. Structural analysis revealed key residues involved in peptide backbone binding including Asp542, Ala543 and Arg717.

Such residues have previously been shown to be critical to inhibitor binding in NEP. Indeed, comparison with other zinc metalloproteases has further illustrated the similar function and importance of these residues in peptide binding.

Substantial steps towards comprehensive characterisation of NEP have been achieved from the work reported in this thesis. There are however a number of points that should be addressed and several avenues for further research. First, characterisation of peptide side chain binding. One of the major limitations of the NEP peptide-bound structure included in Chapter 5 was the inability to model side chain residues for the substrate. This was likely a result of active site residues binding to different regions of the peptide resulting in mixed side chain occupancy in the calculated electron density. To address this, it may be possible to use smaller substrates (3-4 amino acids) to reduce the likelihood of alternative binding conformations.

Second is the investigation into substrate-protein binding interactions proximal to the active site. For this, structures with longer substrates bound would provide information regarding proximal interactions that may be present within the cavity of NEP. Limited success has been achieved so far with large substrate binding so it may be necessary to remove the zinc ion to remove any low-level activity that remains in the NEP E584D mutant.

Finally, conformational fexibility in NEP is a clear area for future research. While further SAXS analysis may provide additional insights into subdomain motion, it may be necessary recruit higher resolution techniques, such as cryo-EM, to determine the exact nature of conformational movement in NEP. Such research may help provide comprehensive characterisation of NEP and indicate why the enzyme exhibits specificity exclusively for smaller substrates.

7.1.2 ECE-1

ECE-1 is another highly potent ADE (Leissring, 2016), but has received little structural and functional characterisation. To date, a single crystal structure of ECE-1 has been published in an inhibitor-bound form (Figure 7.2) (Schulz *et al.*, 2009). The protein used to determine this structure was expressed in *E. coli*, required refolding, and included point mutation C428S to prevent dimerization. No kinetic characterisation had been conducted for this protein so the full effect of these modifications on protein folding and activity are unclear. Despite this,

the structure is validated by the high structural homology to NEP and serves as a good starting point for ECE-1 characterisation.

Chapter 6 outlines the research conducted on ECE-1. Similar to NEP, characterisation of ECE-1 first required the development of a method for protein expression and purification. Initial expression trials were performed using the protocol described in (Schulz *et al.,* 2009). However, limited information regarding the method of protein refolding prevented the successful replication of this protocol and the production of correctly folded protein was not achieved.

Additional fungal and mammalian expression systems were used to assess the optimal method for protein expression and purification. While soluble protein was generated in both expression systems, the ECE-1 yield was low. Despite this, the results included in Chapter 6 suggest that the *P. pastoris* expression system was the most appropriate for ECE-1 expression and, given more time, it may be possible to improve protein yield. Without the required quantity of protein required for crystallisation trials the work on ECE-1 did not progress into structural determination. Despite this, structural analysis was possible through comparison of the available ECE-1 structure with the NEP peptide-bound structure detailed in Chapter 5.

Future research into ECE-1 could follow the same narrative of NEP. Determination of peptide-bound structures would provide information regarding how specificity is achieved. Likewise, research into the conformational flexibility in ECE-1 could be achieved through SAXS or cryo-EM analysis. Such methods, if used on dimeric ECE-1, would also provide information regarding the spatial organisation of subunits.

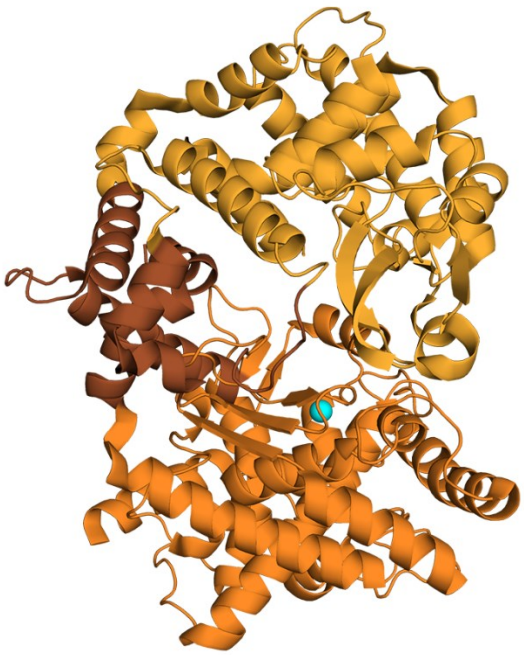


Figure 7.2, Crystal structure of the extracellular domain of ECE-1 in an inhibitor-bound form (3DWB). The protein contains a large central cavity where the active site is located based around a zinc ion (cyan).

7.2 Concluding remarks

The field of Alzheimer's disease research is one that encompasses a variety of scientific disciplines, principally neuroscience, biochemistry and pharmacology. The increased focus placed on Alzheimer's disease, particularly over the last twenty years, has led to a vast collection of literature surrounding the field. To date, some 181,116 articles have been published (Dong *et al.,* 2019), covering a broad range of topics including technological advances in disease diagnosis, identification of genetic risk factors and molecular biomarkers associated with disease initiation and progression, protein networks involved in possible molecular mechanisms, while also including interdisciplinary investigations such as the role of inflammation in disease progression.

While this information has helped to identify individuals with higher risk of developing Alzheimer's disease, and provide disease preventing lifestyle guidelines (Barnard *et al.,* 2014), current research has yet to yield therapeutic agents capable of halting or significantly reducing disease progression. A major limitation to this is the absence of a concrete molecular mechanism explaining how Alzheimer's disease develops.

Presently a large quantity of evidence, discussed throughout the introduction of this thesis, firmly points to the peptide amyloid-β as playing a critical role in disease initiation and progression. Indeed, the recent positive clinical trial results for amyloid-β antibody therapy – aducanumab – has reinforced the importance of amyloid-β in Alzheimer's disease research.

If therapies targeting amyloid-β can yield reductions in disease progression, it is clear that a significant focus should be placed on mechanisms governing amyloid-β regulation. Of these mechanisms ADE's represent key means of amyloid-β regulation, and, while a strong foundation of knowledge has been accumulated full characterisation of these proteins has yet to be achieved. Because of this, ADE's will remain a key area of the Alzheimer's disease research.

References

Adams, P.D., Afonine, P. V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. and Zwart, P.H.,(2010). Phenix: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*, **66**(2), pp.213–221.

Afonine, P. V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., IUCr, M., R.L., T., S. and L., W.G.,(2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D Biological Crystallography*, **68**(4), pp.352–367.

Ahmed, M., Davis, J., Aucoin, D., Sato, T., Ahuja, S., Aimoto, S., Elliott, J.I., Van Nostrand, W.E. and Smith, S.O.,(2010). Structural conversion of neurotoxic amyloid-β1–42 oligomers to fibrils. *Nature Structural & Molecular Biology*, **17**(5), pp.561–567.

Almenoff, J., Wilk, S. and Orlowski, M.,(1981). Membrane bound pituitary metalloendopeptidase: Apparent identity to enkephalinase. *Biochemical and Biophysical Research Communications*, **102**(1), pp.206–214.

Backstrom, J.R., Lim, G.P., Cullen, M.J. and Tökés, Z.A.,(1996). Matrix metalloproteinase-9 (MMP-9) is synthesized in neurons of the human hippocampus and is capable of degrading the amyloid-β peptide (1-40). *Journal of Neuroscience*, **16**(24), pp.7910–7919.

Bailey, J.A., Maloney, B., Ge, Y.-W. and Lahiri, D.K.,(2011). Functional activity of the novel Alzheimer's amyloid β-peptide interacting domain (AβID) in the APP and BACE1 promoter sequences and implications in activating apoptotic genes and in amyloidogenesis. *Gene*, **488**(1–2), pp.13–22.

Baker-Nigh, A., Vahedi, S., Davis, E.G., Weintraub, S., Bigio, E.H., Klein, W.L. and Geula, C.,(2015). Neuronal amyloid-β accumulation within cholinergic basal forebrain in ageing and Alzheimer's disease. *Brain*, **138**(6), pp.1722–1737.

Barnard, N.D., Bush, A.I., Ceccarelli, A., Cooper, J., de Jager, C.A., Erickson, K.I., Fraser, G., Kesler, S., Levin, S.M., Lucey, B., Morris, M.C. and Squitti, R.,(2014). Dietary and lifestyle guidelines for the prevention of Alzheimer's disease. *Neurobiology of Aging*, **35**(S2), pp.S74–S78.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R. and Leslie, A.G.W.,(2011). *iMOSFLM*: a new graphical interface for diffraction-image processing with *MOSFLM*. *Acta Crystallographica Section D Biological Crystallography*, **67**(4), pp.271–281.

Bayes-Genis, A., Barallat, J. and Richards, A.M.,(2016). A test in context: neprilysin: function, inhibition, and biomarker. *Journal of the American College of Cardiology*, **68**(6), pp.639–653.

Beher, D., Hesse, L., Masters, C.L. and Multhaup, G.,(1996). Regulation of amyloid protein precursor (APP) binding to collagen and mapping of the binding sites on APP and collagen type I. *The Journal of Biological Chemistry*, **271**(3), pp.1613–1620.

Berman, H., Henrick, K. and Nakamura, H.,(2003). Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology*, **10**(12), p.980.

Bertram, L., McQueen, M.B., Mullin, K., Blacker, D. and Tanzi, R.E.,(2007). Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. *Nature Genetics*, **39**(1), pp.17–23.

Bibby, J., Keegan, R.M., Mayans, O., Winn, M.D. and Rigden, D.J.,(2012). *AMPLE*: a cluster-and-truncate approach to solve the crystal structures of small proteins using rapidly computed *ab initio* models. *Acta Crystallographica Section D Biological Crystallography*, **68**(12), pp.1622–1631.

Bird, T.D.,(1999). *Early-Onset Familial Alzheimer Disease*. In: M.P, Adam. Gene Reviews. Seattle: University of Washington. Available from: http://www.ncbi.nlm.nih.gov/pubmed/ 20301414 [Accessed 2019].

Bitan, G., Kirkitadze, M.D., Lomakin, A., Vollers, S.S., Benedek, G.B. and Teplow, D.B.,(2003). Amyloid β protein (Aβ) assembly: Aβ 40 and Aβ 42 oligomerize through distinct pathways. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(1), pp.330–335.

Blow, D.M.,(2004). *Outline of crystallography for biologists*. Oxford: Oxford University Press.

Boldon, L., Laliberte, F. and Liu, L.,(2015). Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. *Nano Reviews*, **6**(1). Available from: https://www.tandfonline.com/doi/full/10.3402/nano.v6.25661 [Accessed 2019].

Bolduc, D.M., Montagna, D.R., Seghers, M.C., Wolfe, M.S., Selkoe, D.J., Selkoe, D., Ihara, Y., Koo, E., Haass, C., Takio, K., Morishima-Kawashima, M., Ishiura, S., Ihara, Y., Smith, A., Shi, X., Yin, K., Shafer, J., Gardell, S., Pastor, P., Pérez-Tur, J., Relvas, R., Oliveira, C., Ribeiro, M., Rogaeva, E., Sa, A., Samaranch, L., Sánchez-Valle, R., Santana, I., Tàrraga, L., Valdivieso, F., Singleton, A., Hardy, J. and Clarimón, J.,(2016). The amyloid-β forming tripeptide cleavage mechanism of γ-secretase. *eLife*, **5**. Available from: https://elifesciences.org/articles/17578 [Accessed 2019].

Bondy, C.A., Zhou, J., Chin, E., Reinhardt, R.R., Ding, L. and Roth, R.A.,(1994). Cellular distribution of insulin-degrading enzyme gene expression. Comparison with insulin and insulin-like growth factor receptors. *The Journal of Clinical Investigation*, **93**(3), pp.966–973.

Bragg, W.H. and Bragg, W.L.,(1913). The reflection of X-rays by crystals. (II.). *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **89**(610), pp.246–248.

Brill, R.,(1968). Determination of electron distribution in crystals by means of X-rays. In: S, Federick., D, Turnball., H, Ehrenreich. *Solid State Physics*. Academic Press, pp.1–35.

Bünning, P. and Riordan, J.F.,(1985). The functional role of zinc in angiotensin converting enzyme: implications for the enzyme mechanism. *Journal of Inorganic Biochemistry*, **24**(3), pp.183–198.

Campbell, D.J.,(2017). Long-term neprilysin inhibition — implications for ARNIs. *Nature Reviews Cardiology*, **14**(3), pp.171–186.

Campion, D., Dumanchin, C., Hannequin, D., Dubois, B., Belliard, S., Puel, M., Thomas-Anterion, C., Michon, A., Martin, C., Charbonnier, F., Raux, G., Camuzat, A., Penet, C., Mesnage, V., Martinez, M., Clerget-Darpoux, F., Brice, A. and Frebourg, T.,(1999). Early-

onset autosomal dominant Alzheimer disease: Prevalence, genetic heterogeneity, and mutation spectrum. *American Journal of Human Genetics*, **65**(3), pp.664–670.

Cerdà-Costa, N. and Gomis-Rüth, F.X.,(2014). Architecture and function of metallopeptidase catalytic domains. *Protein Science*, **23**(2), pp.123–144.

Chang, V.T., Crispin, M., Aricescu, A.R., Harvey, D.J., Nettleship, J.E., Fennelly, J.A., Yu, C., Boles, K.S., Evans, E.J., Stuart, D.I., Dwek, R.A., Jones, E.Y., Owens, R.J. and Davis, S.J.,(2007). Glycoprotein structural genomics: solving the glycosylation problem. *Structure*, **15**(3), pp.267–73.

Chen, G., Xu, T., Yan, Y., Zhou, Y., Jiang, Y., Melcher, K. and Xu, H.E.,(2017). Amyloid beta: structure, biology and structure-based therapeutic development. *Acta Pharmacologica Sinica*, **38**(9), pp.1205–1235.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C.,(2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica Section D Biological Crystallography*, **66**(1), pp.12–21.

Cheng, I.H., Scearce-Levie, K., Legleiter, J., Palop, J.J., Gerstein, H., Bien-Ly, N., Puoliväli, J., Lesné, S., Ashe, K.H., Muchowski, P.J. and Mucke, L.,(2007). Accelerating amyloid-β fibrillization reduces oligomer levels and functional deficits in Alzheimer disease mouse models. *Journal of Biological Chemistry*, **282**(33), pp.23818–23828.

Citron, M., Oltersdorf, T., Haass, C., McConlogue, L., Hung, A.Y., Seubert, P., Vigo-Pelfrey, C., Lieberburg, I. and Selkoe, D.J.,(1992). Mutation of the β-amyloid precursor protein in familial Alzheimer's disease increases β-protein production. *Nature*, **360**(6405), pp.672–674.

Clarris, H.J., Cappai, R., Heffernan, D., Beyreuther, K., Masters, C.L. and Small, D.H.,(1997). Identification of heparin-binding domains in the amyloid precursor protein of Alzheimer's disease by deletion mutagenesis and peptide mapping. *Journal of Neurochemistry*, **68**(3), pp.1164–1172.

Coles, M., Bicknell, W., Watson, A.A., Fairlie, D.P. and Craik, D.J.,(1998). Solution structure of amyloid beta-peptide(1-40) in a water-micelle environment. Is the membrane-

spanning domain where we think it is? *Biochemistry*, **37**(31), pp.11064–11077.

Colvin, M.T., Silvers, R., Ni, Q.., Can, T. V, Sergeyev, I., Rosay, M., Donovan, K.J., Michael, B., Wall, J., Linse, S. and Griffin, R.G.,(2016). Atomic resolution structure of monomorphic Aβ 42 amyloid fibrils. *Journal of the American Chemical Society*, **138**(30), pp.9663–9674.

Cook, D.G., Forman, M.S., Sung, J.C., Leight, S., Kolson, D.L., Iwatsubo, T., Lee, V.M.-Y. and Doms, R.W.,(1997). Alzheimer's Aβ(1–42) is generated in the endoplasmic reticulum/intermediate compartment of NT2N cells. *Nature Medicine*, **3**(9), pp.1021–1023.

Corradi, H.R., Schwager, S.L.U., Nchinda, A.T., Sturrock, E.D. and Acharya, K.R.,(2006). Crystal structure of the N domain of human somatic angiotensin I-converting enzyme provides a structural basis for domain-specific inhibitor design. *Journal of Molecular Biology*, **357**(3), pp.964–974.

Cowtan, K.,(2003). Phase problem in X-ray crystallography, and its solution. *Encyclopedia of life sciences*. Available from https://onlinelibrary.wiley.com/doi/full/10.1038/npg.els.0002722 [Accessed 2019].

Cowtan, K.,(2006). The *Buccaneer* software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D Biological Crystallography*, **62**(9), pp.1002–1011.

Cregg, J.M., Vedvick, T.S. and Raschke, W.C.,(1993). Recent advances in the expression of foreign genes in *Pichia pastoris*. *Biotechnology*, **11**(8), pp.905–910.

Crescenzi, O., Tomaselli, S., Guerrini, R., Salvadori, S., D'Ursi, A.M., Temussi, P.A. and Picone, D.,(2002). Solution structure of the Alzheimer amyloid beta-peptide (1-42) in an apolar microenvironment. Similarity with a virus fusion domain. *European Journal of Biochemistry*, **269**(22), pp.5642–5648.

Croset, A., Delafosse, L., Gaudry, J.P., Arod, C., Glez, L., Losberger, C., Begue, D., Krstanovic, A., Robert, F., Vilbois, F., Chevalet, L. and Antonsson, B.,(2012). Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells. *Journal of Biotechnology*, **161**(3), pp.336–348.

Dale, G.E., Oefner, C. and D'Arcy, A.,(2003). The protein as a variable in protein crystallization. *Journal of Structural Biology*, **142**(1), pp.88–97.

Dauter, Z.,(2017). Collection of X-Ray diffraction data from macromolecular crystals. *Methods in Molecular Biology*, **1607**, pp.165–184.

Dauter, Z. and Wlodawer, A.,(2016). Progress in protein crystallography. *Protein and Peptide Letters*, **23**(3), pp.201–210.

Davenport, A.P., Kuc, R.E., Plumpton, C., Mockridge, J.W., Barker, P.J. and Huskisson, N.S.,(1998). Endothelin-converting enzyme in human tissues. *The Histochemical Journal*, **30**(5), pp.359–374.

Davis, D.G., Schmitt, F.A., Wekstein, D.R. and Markesbery, W.R.,(1999). Alzheimer neuropathologic alterations in aged cognitively normal subjects. *Journal of Neuropathology and Experimental Neurology*, **58**(4), pp.376–388.

Davis, Simon J, Crispin, Max, Davis, S J and Crispin, M,(2011). Solutions to the glycosylation problem for low-and high-throughput structural glycoproteomics. In: R. Owens and J. Nettleship. *Functional and Structural Proteomics of Glycoproteins*, pp.127–158.

Deane, R., Sagare, A., Hamm, K., Parisi, M., Lane, S., Finn, M.B., Holtzman, D.M. and Zlokovic, B. V,(2008). apoE isoform – specific disruption of amyloid β peptide clearance from mouse brain. *The Journal of Clinical Investigation*, **118**(12), pp.4002–4013.

Deane, R., Du Yan, S., Submamaryan, R.K., LaRue, B., Jovanovic, S., Hogg, E., Welch, D., Manness, L., Lin, C., Yu, J., Zhu, H., Ghiso, J., Frangione, B., Stern, A., Schmidt, A.M., Armstrong, D.L., Arnold, B., Liliensiek, B., Nawroth, P., Hofman, F., Kindy, M., Stern, D. and Zlokovic, B.,(2003). RAGE mediates amyloid-β peptide transport across the blood-brain barrier and accumulation in brain. *Nature Medicine*, **9**(7), pp.907–913.

Dessau, M.A. and Modis, Y.,(2011). Protein crystallization for X-ray crystallography. *Journal of Visualized Experiments*, **16**(47). Available from: https://www.jove.com/t/2285/protein-crystallization-for-x-ray-crystallography [Accessed 2019].

Doggrell, S.A.,(2004). Endothelin-converting enzyme inhibitors and their potential for cardiovascular and renal therapeutics. *Expert Opinion on Therapeutic Patents*, **14**(5),

pp.655–665.

Dong, R., Wang, H., Ye, J., Wang, M. and Bi, Y.,(2019). Publication trends for Alzheimer's disease worldwide and in China: a 30-year bibliometric analysis. *Frontiers in Human Neuroscience*, **13**(259). Available from: https://www.frontiersin.org/articles/ 10.3389/fnhum.2019.00259/full [Acessed 2019].

Doody, R.S., Raman, R., Farlow, M., Iwatsubo, T., Vellas, B., Joffe, S., Kieburtz, K., He, F., Sun, X., Thomas, R.G., Aisen, P.S., Alzheimer's Disease Cooperative Study Steering Committee, Siemers, E., Sethuraman, G., Mohs, R. and Semagacestat Study Group,(2013). A phase 3 trial of semagacestat for treatment of Alzheimer's disease. *The New England Journal of Medicine*, **369**(4), pp.341–350.

Duckworth, W.C., Heinemann, M.A. and Kitabchi, A.E.,(1972). Purification of insulin-specific protease by affinity chromatography. *Proceedings of the National Academy of Sciences of the United States of America*, **69**(12), pp.3698–3702.

van Dyck, C.H.,(2018). Anti-amyloid-β monoclonal antibodies for Alzheimer's disease: pitfalls and promise. *Biological Psychiatry*, **83**(4), pp.311–319.

Dyrks, T., Weidemann, A., Multhaup, G., Salbaum, J.M., Lemaire, H.G., Kang, J., Müller-Hill, B., Masters, C.L. and Beyreuther, K.,(1988). Identification, transmembrane orientation and biogenesis of the amyloid A4 precursor of Alzheimer's disease. *The EMBO journal*, **7**(4), pp.949–957.

Eckman, E.A., Adams, S.K., Troendle, F.J., Stodola, B.A., Kahn, M.A., Fauq, A.H., Xiao, H.D., Bernstein, K.E. and Eckman, C.B.,(2006). Regulation of steady-state-amyloid levels in the brain by neprilysin and endothelin-converting enzyme but not angiotensin-converting enzyme. *The Journal of Biological Chemistry*, **281**(41), pp.30471–30478.

Eckman, E.A., Reed, D.K. and Eckman, C.B.,(2001). Degradation of the Alzheimer's amyloid beta peptide by endothelin-converting enzyme. *The Journal of Biological Chemistry*, **276**(27), pp.24540–24548.

Eckman, E.A., Watson, M., Marlow, L., Sambamurti, K. and Eckman, C.B.,(2003). Alzheimer's disease β-amyloid peptide is increased in mice deficient in endothelin-converting enzyme. *The Journal of Biological Chemistry*, **278**(4), pp.2081–2084.

El-Amouri, S.S., Zhu, H., Yu, J., Marr, R., Verma, I.M. and Kindy, M.S.,(2008). Neprilysin: an enzyme candidate to slow the progression of Alzheimer's disease. *The American Journal of Pathology*, **172**(5), pp.1342-1354.

Ellis, S.B., Brust, P.F., Koutz, P.J., Waters, A.F., Harpold, M.M. and Gingeras, T.R.,(1985). Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. *Molecular and Cellular Biology*, **5**(5), pp.1111–1121.

Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K.,(2010). Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography*, **66**(4), pp.486–501.

Engh, R.A., Huber, R. and IUCr,(1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallographica Section A Foundations of Crystallography*, **47**(4), pp.392–400.

Erdös, E.G. and Skidgel, R.A.,(1989). Neutral endopeptidase 24.11 (enkephalinase) and related regulators of peptide hormones. *FASEB Journal*, **3**(2), pp.145–151.

Evans, G. and Pettifer, R.F.,(2001). *CHOOCH*: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *Journal of Applied Crystallography*, **34**(1), pp.82–86.

Evans, P.,(2006). Scaling and assessment of data quality. *Acta Crystallographica Section D Biological Crystallography*, **62**(1), pp.72–82.

Evans, P. and McCoy, A.,(2008). An introduction to molecular replacement. *Acta Crystallographica Section D Biological Crystallography*, **64**(1), pp.1–10.

Evans, P.R. and Murshudov, G.N.,(2013). How good are my data and what is the resolution? *Acta Crystallographica Section D Biological Crystallography*, **69**(7), pp.1204–1214.

Fagan, A.M., Mintun, M.A., Shah, A.R., Aldea, P., Roe, C.M., Mach, R.H., Marcus, D., Morris, J.C. and Holtzman, D.M.,(2009). Cerebrospinal fluid tau and ptau181 increase with cortical amyloid deposition in cognitively normal individuals: Implications for future clinical trials of Alzheimer's disease. *EMBO Molecular Medicine*, **1**(8–9), pp.371–380.

Fan, X., Wang, J., Zhang, X., Yang, Z., Zhang, J.-C., Zhao, L., Peng, H.-L., Lei, J. and Wang, H.-W.,(2019). Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2

Angstrom resolution. *Nature Communications*, **10**(1). Available from: https://www.nature.com/articles/s41467-019-10368-w [Accessed 2019].

Farris, W., Mansourian, S., Chang, Y., Lindsley, L., Eckman, E.A., Frosch, M.P., Eckman, C.B., Tanzi, R.E., Selkoe, D.J. and Guenette, S.,(2003). Insulin-degrading enzyme regulates the levels of insulin, amyloid beta-protein, and the beta-amyloid precursor protein intracellular domain *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(7), pp.4162–4167.

Farris, W., Schütz, S.G., Cirrito, J.R., Shankar, G.M., Sun, X., George, A., Leissring, M.A., Walsh, D.M., Qiu, W.Q., Holtzman, D.M. and Selkoe, D.J.,(2007). Loss of neprilysin function promotes amyloid plaque formation and causes cerebral amyloid angiopathy. *The American Journal of Pathology*, **171**(1), pp.241–251.

De Felice, F.G., Wu, D., Lambert, M.P., Fernandez, S.J., Velasco, P.T., Lacor, P.N., Bigio, E.H., Jerecic, J., Acton, P.J., Shughrue, P.J., Chen-Dodson, E., Kinney, G.G. and Klein, W.L.,(2007). Alzheimer's disease-type neuronal tau hyperphosphorylation induced by Aβ oligomers. *Neurobiology of Aging*, **29**(9), pp.1334–1347.

Fodje, M., Grochulski, P., Janzen, K., Labiuk, S., Gorin, J. and Berg, R.,(2014). 08B1-1: an automated beamline for macromolecular crystallography experiments at the Canadian Light Source. *Journal of Synchrotron Radiation*, **21**(3), pp.633–637.

Franke, D., Petoukhov, M. V, Konarev, P. V, Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., Svergun, D.I. and Semenov, N.N.,(2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography*, **50**(4), pp.1212–1225.

Frautschy, S.A., Yang, F., Irrizarry, M., Hyman, B., Saido, T.C., Hsiao, K. and Cole, G.M.,(1998). Microglial response to amyloid plaques in APPsw transgenic mice. *The American Journal of Pathology*, **152**(1), pp.307–317.

Fukami, S., Watanabe, K., Iwata, N., Haraoka, J., Lu, B., Gerard, N.P., Gerard, C., Fraser, P., Westaway, D., George-Hyslop, P. St. and Saido, T.C.,(2002). Aβ-degrading endopeptidase, neprilysin, in mouse brain: synaptic and axonal localization inversely

correlating with Aβ pathology. *Neuroscience Research*, **43**(1), pp.39–56.

Gafford, J.T., Skidgel, R.A., Erdos, E.G. and Hersh, L.B.,(1983). Human kidney 'enkephalinase', a neutral metalloendopeptidase that cleaves active peptides. *Biochemistry*, **22**(13), pp.3265–3271.

Gehm, B.D., Kuo, W.L., Perlman, R.K. and Rosner, M.R.,(1993). Mutations in a zinc-binding domain of human insulin-degrading enzyme eliminate catalytic activity but not insulin binding. *The Journal of Biological Chemistry*, **268**(11), pp.7943–7948.

Glenner, G.G. and Wong, C.W.,(1984). Alzheimer's disease and Down's syndrome: Sharing of a unique cerebrovascular amyloid fibril protein. *Biochemical and Biophysical Research Communications*, **122**(3), pp.1131–1135.

Glossop, M.S., Bazin, R.J., Dack, K.N., Fox, D.N.A., MacDonald, G.A., Mills, M., Owen, D.R., Phillips, C., Reeves, K.A., Ringer, T.J., Strang, R.S. and Watson, C.A.L.,(2011). Synthesis and evaluation of heteroarylalanine diacids as potent and selective neutral endopeptidase inhibitors. *Bioorganic & Medicinal Chemistry Letters*, **21**(11), pp.3404–3406.

Goate, A., Chartier-Harlin, M.-C., Mullan, M., Brown, J., Crawford, F., Fidani, L., Giuffra, L., Haynes, A., Irving, N., James, L., Mant, R., Newton, P., Rooke, K., Roques, P., Talbot, C., Pericak-Vance, M., Roses, A., Williamson, R., Rossor, M., Owen, M. and Hardy, J.,(1991). Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature*, **349**(6311), pp.704–706.

Goldstein, L.S.B.,(2012). Axonal transport and neurodegenerative disease: Can we see the elephant? *Progress in Neurobiology*, **99**(3), pp.186–190.

Gomis-Rüth, F.,(2008). Structure and mechanism of metallocarboxypeptidases. *Critical Reviews in Biochemistry and Molecular Biology*, **43**(5), pp.319–345.

Gong, Y., Chang, L., Viola, K.L., Lacor, P.N., Lambert, M.P., Finch, C.E., Krafft, G.A. and Klein, W.L.,(2003). Alzheimer's disease-affected brain: presence of oligomeric A beta ligands (ADDLs) suggests a molecular basis for reversible memory loss. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(18), pp.10417–10422.

Graham, W.V., Bonito-Oliva, A. and Sakmar, T.P.,(2017). Update on Alzheimer's disease therapy and prevention strategies. *Annual Review of Medicine*, **68**(1), pp.413–430.

Gremer, L., Schölzel, D., Schenk, C., Reinartz, E., Labahn, J., Ravelli, R.B.G., Tusche, M., Lopez-Iglesias, C., Hoyer, W., Heise, H., Willbold, D. and Schröder, G.F.,(2017). Fibril structure of amyloid-β(1–42) by cryo–electron microscopy. *Science*, **358**(6359), pp.116–119.

Grinna, L.S. and Tschopp, J.F.,(1989). Size distribution and general structural features ofN-linked oligosaccharides from the methylotrophic yeast *Pichia pastoris*. *Yeast*, **5**(2), pp.107–115.

Haass, C., Kaether, C., Thinakaran, G. and Sisodia, S.,(2012). Trafficking and proteolytic processing of APP. *Cold Spring Harbor Perspectives in Medicine*, **2**(5). Available from: http://perspectivesinmedicine.cshlp.org/content/2/5/a006270 [Acessed 2019].

Haass, C., Koo, E.H., Mellon, A., Hung, A.Y. and Selkoe, D.J.,(1992). Targeting of cell-surface β-amyloid precursor protein to lysosomes: alternative processing into amyloid-bearing fragments. *Nature*, **357**(6378), pp.500–503.

Hardy, J.,(1997). Amyloid, the presenilins and Alzheimer's disease. *Trends in Neurosciences*, **20**(4), pp.154–159.

Hardy, J.A. and Higgins, G.A.,(1992). Alzheimer's disease: the amyloid cascade hypothesis. *Science*, **256**(5054), pp.184–185.

Hartmann, T., Bieger, S.C., Brühl, B., Tienari, P.J., Ida, N., Allsop, D., Roberts, G.W., Masters, C.L., Dotti, C.G., Unsicker, K. and Beyreuther, K.,(1997). Distinct sites of intracellular production for Alzheimer's disease Aβ40/42 amyloid peptides. *Nature Medicine*, **3**(9), pp.1016–1020.

Hartopo, A.B., Emoto, N., Vignon-Zellweger, N., Suzuki, Y., Yagi, K., Nakayama, K. and Hirata, K.,(2013). Endothelin-converting enzyme–1 gene ablation attenuates pulmonary fibrosis via CGRP-cAMP/EPAC1 pathway. *American Journal of Respiratory Cell and Molecular Biology*, **48**(4), pp.465–476.

Hatami, A., Monjazeb, S., Milton, S. and Glabe, C.G.,(2017). Familial Alzheimer's disease mutations within the amyloid precursor protein alter the aggregation and

conformation of the amyloid-β peptide. *Journal of Biological Chemistry*, **292**(8), pp.3172–3185.

Haynes, W.G., Strachan, F.E., Gray, G.A. and Webb, D.J.,(1995). Forearm vasoconstriction to endothelin-1 is mediated by ETA and ETB receptors *in vivo* in humans. *Journal of Cardiovascular Pharmacology*, **26**(S3), pp.40–43.

Head, E., Powell, D., Gold, B. and Schmutt, F.,(2012). Alzheimer's disease in Down's syndrome. *European Journal of Neurodegenerative Disease*, **75**(5), pp.353–364.

Helliwell, J.R.,(2017). New developments in crystallography: exploring its technology, methods and scope in the molecular biosciences. *Bioscience Reports*, **37**(4). Available from: https://portlandpress.com/bioscirep/article/37/4/BSR20170204/57104/New-developments-in-crystallography-exploring-its [Acessed 2019].

Hemming, M.L. and Selkoe, D.J.,(2005). Amyloid-protein is degraded by cellular angiotensin-converting enzyme (ACE) and elevated by an ACE Inhibitor. *The Journal of Biological Chemistry*, **280**(45), pp.37644–37650.

Hendrickson, W.,(1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**(5028), pp.51–58.

Henley, D.B., Sundell, K.L., Sethuraman, G., Dowsett, S.A. and May, P.C.,(2014). Safety profile of semagacestat, a gamma-secretase inhibitor: IDENTITY trial findings. *Current Medical Research and Opinion*, **30**(10), pp.2021–2032.

Holmquist, B. and Vallee, B.L.,(1974). Metal substitutions and inhibition of thermolysin: spectra of the cobalt enzyme. *The Journal of Biological Chemistry*, **249**(14), pp.4601–4607.

Hong, C.-S., Goins, W.F., Goss, J.R., Burton, E.A. and Glorioso, J.C.,(2006). Herpes simplex virus RNAi and neprilysin gene transfer vectors reduce accumulation of Alzheimer's disease-related amyloid-beta peptide *in vivo*. *Gene Therapy*, **13**(14), pp.1068–1079.

Howard, T.E., Shai, S.Y., Langford, K.G., Martin, B.M. and Bernstein, K.E.,(1990). Transcription of testicular angiotensin-converting enzyme (ACE) is initiated within the 12th intron of the somatic ACE gene. *Molecular and Cellular Biology*, **10**(8), pp.4294–4302.

Howell, S., Nalbantoglu, J. and Crine, P.,(1995). Neutral endopeptidase can hydrolyze β-amyloid(1-40) but shows no effect on beta-amyloid precursor protein metabolism. *Peptides*, **16**(4), pp.647–652.

Hsia, A.Y., Masliah, E., McConlogue, L., Yu, G.Q., Tatsuno, G., Hu, K., Kholodenko, D., Malenka, R.C., Nicoll, R.A. and Mucke, L.,(1999). Plaque-independent disruption of neural circuits in Alzheimer's disease mouse models. *Proceedings of the National Academy of Sciences of the United States of America*, **96**(6), pp.3228–3233.

Hsiao, K.K., Borchelt, D.R., Olson, K., Johannsdottir, R., Kitt, C., Yunis, W., Xu, S., Eckman, C., Younkin, S., Price, D., Overbeek, P.A., Clark, H.B. and Carlson, G.,(1995). Age-related CNS disorder and early death in transgenic FVB/N mice overexpressing Alzheimer amyloid precursor proteins. *Neuron*, **15**(5), pp.1203–1218.

Hu, J., Igarashi, A., Kamata, M. and Nakagawa, H.,(2001). Angiotensin-converting enzyme degrades alzheimer amyloid-β peptide; retards aggregation, deposition, fibril formation; and inhibits cytotoxicity. *Journal of Biological Chemistry*, **276**(51), pp.47863–47868.

Hu, J., Miyatake, F., Aizu, Y., Nakagawa, H., Nakamura, S., Tamaoka, A., Takahash, R., Urakami, K. and Shoji, M.,(1999). Angiotensin-converting enzyme genotype is associated with Alzheimer disease in the Japanese population. *Neuroscience Letters*, **277**(1), pp.65–67.

Huang, J.Y., Bruno, A.M., Patel, C.A., Huynh, A.M., Philibert, K.D., Glucksman, M.J. and Marr, R.A.,(2008). Human membrane metallo-endopeptidase-like protein degrades both beta-amyloid 42 and beta-amyloid 40. *Neuroscience*, **155**(1), pp.258–262.

Huang, S.-M., Mouri, A., Kokubo, H., Nakajima, R., Suemoto, T., Higuchi, M., Staufenbiel, M., Noda, Y., Yamaguchi, H., Nabeshima, T., Saido, T.C. and Iwata, N.,(2006). Neprilysin-sensitive synapse-associated amyloid-β peptide oligomers impair neuronal plasticity and cognitive function. *Journal of Biological Chemistry*, **281**(26), pp.17941–17951.

Igic, R. and Behnia, R.,(2003). Properties and distribution of angiotensin I converting enzyme. *Current Pharmaceutical Design*, **9**(9), pp.697–706.

Ignjacev-Lazich, I., Kintsurashvili, E., Johns, C., Vitseva, O., Duka, A., Shenouda, S., Gavras, I. and Gavras, H.,(2005). Angiotensin-converting enzyme regulates bradykinin receptor

gene expression. *American Journal of Physiology - Heart and Circulatory Physiology*, **289**(5), pp.H1814–H1820.

Incardona, M.-F., Bourenkov, G.P., Levik, K., Pieritz, R.A., Popov, A.N. and Svensson, O.,(2009). *EDNA*: a framework for plugin-based applications applied to X-ray experiment online data analysis. *Journal of Synchrotron Radiation*, **16**(6), pp.872–879.

Iwata, N., Tsubuki, S., Takaki, Y., Shirotani, K., Lu, B., Gerard, N.P., Gerard, C., Hama, E., Lee, H.J. and Saido, T.C.,(2001). Metabolic regulation of brain Aβ by neprilysin. *Science*, **292**(5521), pp.1550–1552.

Jacobsen, J.S., Comery, T.A., Martone, R.L., Elokdah, H., Crandall, D.L., Oganesian, A., Aschmies, S., Kirksey, Y., Gonzales, C., Xu, J., Zhou, H., Atchison, K., Wagner, E., Zaleska, M.M., Das, I., Arias, R.L., Bard, J., Riddell, D., Gardell, S.J., Abou-Gharbia, M., Robichaud, A., Magolda, R., Vlasuk, G.P., Bjornsson, T., Reinhart, P.H. and Pangalos, M.N.,(2008). Enhanced clearance of Aβ in brain by sustaining the plasmin proteolysis cascade. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(25), pp.8754–8759.

Jaskolski, M., Gilski, M., Dauter, Zbigniew and Wlodawer, A.,(2007). Stereochemical restraints research papers. *Acta Crystallographica Section D Biological Crystallography*, **63**(5), pp.611–620.

Jenkins, H.T. and IUCr,(2018). *Fragon*: rapid high-resolution structure determination from ideal protein fragments. *Acta Crystallographica Section D Structural Biology*, **74**(3), pp.205–214.

Jhund, P.S. and McMurray, J.J. V,(2016). The neprilysin pathway in heart failure: a review and guide on the use of sacubitril/valsartan. *Heart*, **102**(17), pp.1342–1347.

Jiang, S., Li, Y., Zhang, X., Bu, G., Xu, H., Zhang, Y., Marino, D. Di, Borroni, B., Tramontano, A., Perez-Otano, I. and Luca, M. Di,(2014). Trafficking regulation of proteins in Alzheimer's disease. *Molecular Neurodegeneration*, **9**(1). Available from: https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/1750-1326-9-6 [Accessed 2019].

Johnson, A.R., Skidgel, Randal A, Gafford, John T, Erdos, Ervin G, Skidgel, R A, Gafford, J T

and Erdos, E G,(1984). Enzymes in placental microvilli: angiotensin I converting enzyme, angiotensinase A, carboxypeptidase, and neutral endopeptidase. *Peptides*, **5**(4), pp.789–796.

Johnson, G.D. and Ahn, K.,(2000). Development of an internally quenched fluorescent substrate selective for endothelin-converting enzyme-1. *Analytical Biochemistry*, **286**(1), pp.112–118.

Johnson, G.D., Stevenson, T. and Ahn, K.,(1999). Hydrolysis of peptide hormones by endothelin-converting enzyme-1. A comparison with neprilysin. *The Journal of Biological Chemistry*, **274**(7), pp.4053–4058.

Johnson, G.D., Swenson, H.R., Ramage, R. and Ahn, K.,(2002). Mapping the active site of endothelin-converting enzyme-1 through subsite specificity and mutagenesis studies: a comparison with neprilysin. *Archives of Biochemistry and Biophysics*, **398**(2), pp.240–248.

Johnson, K.A., Bhushan, S., Ståhl, A., Hallberg, B.M., Frohn, A., Glaser, E. and Eneqvist, T.,(2006). The closed structure of presequence protease PreP forms a unique 10 000 $Å^3$ chamber for proteolysis. *The EMBO Journal*, **25**(9), pp.1977–1986.

Jung Lee, S.C., Nam, E., Jin Lee, H., Savelieff cd, M.G. and Hee Lim, M.,(2017). Towards an understanding of amyloid-β oligomers: characterization, toxicity mechanisms, and inhibitors. *Chemical Society Reviews* **46**(2), pp.310-323.

Kabsch, W.,(2010). XDS. *Acta Crystallographica Section D: Biological Crystallography*, **66**(2), pp.125-132.

Kalk, P., Sharkovska, Y., Kashina, E., Von Websky, K., Relle, K., Pfab, T., Alter, M., Guillaume, P., Provost, D., Hoffmann, K., Fischer, Y. and Hocher, B.,(2011). Endothelin-converting enzyme/neutral endopeptidase inhibitor SLV338 prevents hypertensive cardiac remodeling in a blood pressure-independent manner. *Hypertension*, **57**(4), pp.755–763.

Kang, J., Lemaire, H.-G., Unterbeck, A., Salbaum, J.M., Masters, C.L., Grzeschik, K.-H., Multhaup, G., Beyreuther, K. and Müller-Hill, B.,(1987). The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature*, **325**(6106), pp.733–736.

Van Der Kant, R. and Goldstein, L.S.B.,(2015). Cellular functions of the amyloid precursor protein from development to dementia. *Developmental Cell*, **32**(4), pp.502–515.

Karplus, P.A. and Diederichs, K.,(2012). Linking crystallographic model and data quality. *Science*, **336**(6084), pp.1030–1033.

Karplus, P.A. and Diederichs, K.,(2015). Assessing and maximizing data quality in macromolecular crystallography. *Current Opinion in Structural Biology*, **34**, pp.60–68.

Kehoe, P.G., Russ, C., McIlroy, S., Williams, H., Holmans, P., Holmes, C., Liolitsa, D., Vahidassr, D., Powell, J., McGleenon, B., Liddell, M., Plomin, R., Dynan, K., Williams, N., Neal, J., Cairns, N.J., Wilcock, G., Passmore, P., Lovestone, S., Williams, J. and Owen, M.J.,(1999). Variation in DCP1, encoding ACE, is associated with susceptibility to Alzheimer disease. *Nature Genetics*, **21**(1), pp.71–72.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J.E.,(2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, **10**(6), pp.845–858.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wychoff, H. and Phillips, D.C.,(1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, **181**(4610), pp.662–666.

Kenny, A.J., Bourne, A. and Ingram, J.,(1993). Hydrolysis of human and pig brain natriuretic peptides, urodilatin, C-type natriuretic peptide and some C-receptor ligands by endopeptidase-24.11. *Biochemical Journal*, **291**(1), pp.83–88.

Kibbey, M.C., Jucker, M., Weeks, B.S., Neve, R.L., Van Nostrand, W.E. and Kleinman, H.K.,(1993). beta-Amyloid precursor protein binds to the neurite-promoting IKVAV site of laminin. *Proceedings of the National Academy of Sciences of the United States of America*, **90**(21), pp.10150–10153.

Kikhney, A.G. and Svergun, D.I.,(2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters*, **589**(19), pp.2570–2577.

Knops, J., Suomensaari, S., Lee, M., McConlogue, L., Seubert, P. and Sinha, S.,(1995). Cell-

type and amyloid precursor protein-type specific inhibition of Aβ release by bafilomycin A1, a selective inhibitor of vacuolar ATPases. *Journal of Biological Chemistry*, **270**(6), pp.2419–2422.

Korostyshevskaya, I.M., Maksimov, V.F. and Rudenko, N.S.,(2016). C-Type natriuretic peptide: what, where and why? *Neuroscience and Behavioral Physiology*, **46**(8), pp.888–894.

Korth, P., Bohle, R.M., Corvol, P. and Pinet, F.,(1999). Cellular distribution of endothelin-converting enzyme-1 in human tissues. *Journal of Histochemistry & Cytochemistry*, **47**(4), pp.447–461.

Korth, P., Egidy, G., Parnot, C., Lemoullec, J.M., Corvol, P. and Pinet, F.,(1997). Construction, expression and characterization of a soluble form of human endothelin-converting-enzyme-1. *FEBS Letters*, **417**(3), pp.365–370.

Lacor, P.N., Buniel, M.C., Furlow, P.W., Clemente, A.S., Velasco, P.T., Wood, M., Viola, K.L. and Klein, W.L.,(2007). Neurobiology of disease d oligomer-induced aberrations in synapse composition, shape, and density provide a molecular basis for loss of connectivity in Alzheimer's disease. *The Journal of Neuroscience*, **27**(4), pp.796–807.

Lafrance, M.H., Vézina, C., Wang, Q., Boileau, G., Crine, P. and Lemay, G.,(1994). Role of glycosylation in transport and enzymic activity of neutral endopeptidase-24.11. *The Biochemical Journal*, **302**, pp.451–4.

Lambert, M.P., Barlow, A.K., Chromy, B.A., Edwards, C., Freed, R., Liosatos, M., Morgan, T.E., Rozovsky, I., Trommer, B., Viola, K.L., Wals, P., Zhang, C., Finch, C.E., Krafft, G.A. and Klein, W.L.,(1998). Diffusible, nonfibrillar ligands derived from Aβ 1-42 are potent central nervous system neurotoxins. *The Journal of Neuroscience* **95**(11), pp.6448–6453.

Lambert, M.P., Viola, K.L., Chromy, B.A., Chang, L., Morgan, T.E., Yu, J., Venton, D.L., Krafft, G.A., Finch, C.E. and Klein, W.L.,(2001). Vaccination with soluble Aβ oligomers generates toxicity-neutralizing antibodies. *Journal of Neurochemistry*, **79**(3), pp.595–605.

Larmuth, K.M., Masuyer, G., Douglas, R.G., Schwager, S.L., Acharya, K Ravi, Sturrock, E.D., Acharya, K R and Sturrock, E.,(2014). Kinetic and structural characterization of amyloid-β peptide hydrolysis by human angiotensin-1-converting enzyme. *FEBS Journal*, **283**(6), pp.1060–1076.

Leissring, M.A.,(2008). The AβCs of Aβ-cleaving proteases. *Journal of Biological Chemistry*, **283**(44), pp.29645–29649.

Leissring, M.A.,(2016). Aβ-degrading proteases: therapeutic potential in Alzheimer disease. *CNS Drugs*, **30**(8), pp.667–675.

Leissring, M.A., Farris, W., Chang, A.Y., Walsh, D.M., Wu, X., Sun, X., Frosch, M.P. and Selkoe, D.J.,(2003). Enhanced proteolysis of β-amyloid in APP transgenic mice prevents plaque formation, secondary pathology, and premature death. *Neuron*, **40**(6), pp.1087–1093.

Lesné, S., Kotilinek, L. and Ashe, K.H.,(2008). Plaque-bearing mice with reduced levels of oligomeric amyloid-β assemblies have intact memory function. *Neuroscience*, **151**(3), pp.745–749.

Li, Y., Wang, J., Zhang, S. and Liu, Z.,(2015). Neprilysin gene transfer: A promising therapeutic approach for Alzheimer's disease. *Journal of Neuroscience Research*, **93**(9), pp.1325–1329.

Long, F., Vagin, A.A., Young, P., Murshudov, G.N. and IUCr,(2008). *BALBES*: a molecular-replacement pipeline. *Acta Crystallographica Section D Biological Crystallography*, **64**(1), pp.125–132.

Lu, J.-X., Qiang, W., Yau, W.-M., Schwieters, C.D., Meredith, S.C. and Tycko, R.,(2013). Molecular structure of β-amyloid fibrils in Alzheimer's disease brain tissue. *Cell*, **154**(6), pp.1257–1268.

Von Lueder, T.G., Sangaralingham, S.J., Wang, B.H., Kompa, A.R., Atar, D., Burnett, J.C. and Krum, H.,(2013). Renin–Angiotensin blockade combined with natriuretic peptide system augmentation. *Circulation: Heart Failure*, **6**(3), pp.594–605.

Luscher, T.F. and Barton, M.,(2000). Endothelins and endothelin receptor antagonists: Therapeutic considerations for a novel class of cardiovascular drugs. *Circulation*, **102**(S19), pp.2434–2440.

Macours, N., Poels, J., Hens, K., Francis, C. and Huybrechts, R.,(2004). Structure, evolutionary conservation, and functions of angiotensin- and endothelin-converting enzymes. *Internation Review of Cytology*, **239**, pp.47–97.

Malito, E., Hulse, R.E. and Tang, W.-J.,(2008). Amyloid-β degrading cryptidases: insulin degrading enzyme, presequence peptidase, and neprilysin. *Cellular and Molecular Life Sciences*, **65**(16), pp.2574–2585.

Manolopoulou, M., Malito, E., Shen, Y., Zhao, J., Neant-Fery, M., Sun, C.-Y., Meredith, S.C., Sisodia, S.S., Leissring, M.A. and Tang, W.-J.,(2007). Structure of substrate-free human insulin-degrading enzyme (IDE) and biophysical analysis of ATP-induced conformational switch of IDE. *Journal of Biological Chemistry*, **282**(35), pp.25453–25463.

Marr, R. and Hafez, D.,(2014). Amyloid beta and Alzheimer's disease: The role of neprilysin-2 in amyloid-beta clearance. *Frontiers in Aging Neuroscience*, **6**(187), pp.1–7.

Marr, R.A., Rockenstein, E., Mukherjee, A., Kindy, M.S., Hersh, L.B., Gage, F.H., Verma, I.M. and Masliah, E.,(2003). Neprilysin gene transfer reduces human amyloid pathology in transgenic mice. *Journal of Neuroscience*, **23**(6), pp.1992–1996.

Masaki, T., Kimura, S., Yanagisawa, M. and Goto, K.,(1991). Molecular and cellular mechanism of endothelin regulation: Implications for vascular function. *Circulation*, **84**(4), pp.1457–1468.

Matthews, B.W.,(1968). Solvent content of protein crystals. *Journal of Molecular Biology*, **33**(2), pp.491–497.

Mawuenyega, K.G., Sigurdson, W., Ovod, V., Munsell, L., Kasten, T., Morris, J.C., Yarasheski, K.E. and Bateman, R.J.,(2010). Decreased clearance of CNS β-amyloid in Alzheimer's disease. *Science*, **330**(6012), p.1774.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J.,(2007). Phaser crystallographic software. *Journal of Applied Crystallography*, **40**(4), pp.658–674.

McNicholas, S., Potterton, E., Wilson, K.S. and Noble, M.E.M.,(2011). Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D Biological Crystallography*, **67**(4), pp.386–394.

McPherson, A. and Cudney, B.,(2014). Optimization of crystallization conditions for

biological macromolecules. *Acta Crystallographica. Section F, Structural Biology Communications*, **70**(11), pp.1445-1467.

McPherson, A. and Gavira, J.A.,(2014). Introduction to protein crystallization. *Acta crystallographica. Section F, Structural Biology Communications*, **70**(1), pp.2–20.

Meilandt, W.J., Cisse, M., Ho, K., Wu, T., Esposito, L.A., Scearce-Levie, K., Cheng, I.H., Yu, G.-Q. and Mucke, L.,(2009). Neprilysin overexpression inhibits plaque formation but fails to reduce pathogenic Abeta oligomers and associated cognitive deficits in human amyloid precursor protein transgenic mice. *Journal of Neuroscience*, **29**(7), pp.1977–1986.

Mellin, V., Jeng, A.Y., Monteil, C., Renet, S., Henry, J.P., Thuillez, C. and Mulder, P.,(2005). Triple ACE-ECE-NEP inhibition in heart failure: A comparison with ACE and dual ECE-NEP inhibition. *Journal of Cardiovascular Pharmacology*, **46**(3), pp.390–397.

Miyauchi, T. and Masaki, T.,(1999). Pathophysiology of endothelin in the cardiovascular system. *Annual Review of Physiology*, **61**, pp.391–415.

Moss, S., Subramanian, V. and Acharya, K.R.,(2018). High resolution crystal structure of substrate-free human neprilysin. *Journal of Structural Biology*, **204**(1), pp.19–25.

Moss, S., Subramanian, V., & Acharya, K. R. (2019). Crystal structure of peptide-bound neprilysin reveals key binding interactions. *FEBS Letters*, **594**(2), pp.327-336.

Mucke, L., Masliah, E., Yu, G.Q., Mallory, M., Rockenstein, E.M., Tatsuno, G., Hu, K., Kholodenko, D., Johnson-Wood, K. and McConlogue, L.,(2000). High-level neuronal expression of Aβ 1-42 in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *The Journal of Neuroscience*, **20**(11), pp.4050–4058.

Mumford, R.A., Pierzchala, P.A., Strauss, A.W. and Zimmerman, M.,(1981). Purification of a membrane-bound metalloendopeptidase from porcine kidney that degrades peptide hormones. *Proceedings of the National Academy of Sciences of the United States of America*, **78**(11), pp.6623–6627.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn,

M.D., Long, F. and Vagin, A.A.,(2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D Biological Crystallography*, **67**(4), pp.355–367.

Nakayama, K., Emoto, N., Suzuki, Y., Vignon-Zellweger, N., Yagi, K. and Hirata, K.,(2012). Physiological relevance of hydrolysis of atrial natriuretic peptide by endothelin-converting enzyme-1. *The Kobe Journal of Medical Sciences*, **58**(1), pp.12–18.

Nalivaeva, N.N., Beckett, C., Belyaev, N.D. and Turner, A.J.,(2012). Are amyloid-degrading enzymes viable therapeutic targets in Alzheimer's disease? *Journal of Neurochemistry*, **120**(S1), pp.167–185.

Nalivaeva, N.N., Belyaev, N.D., Kerridge, C. and Turner, A.J.,(2014). Amyloid-clearing proteins and their epigenetic regulation as a therapeutic target in Alzheimer's disease. *Frontiers in Aging Neuroscience*, **6**, p.235.

Nalivaeva, N.N. and Turner, A.J.,(2013). Neprilysin. *Handbook of Proteolytic Enzymes*. pp.612–619. 7th ed. Academic Press. Available from: https://www.elsevier.com/books/handbook-of-proteolytic-enzymes/ [Accessed Novmember 2019]

Nalivaeva, N.N. and Turner, A.J.,(2019). Targeting amyloid clearance in Alzheimer's disease as a therapeutic strategy. *British Journal of Pharmacology*, **176**(18), pp.3447–3463.

Náray-Szabó, G.,(1993). Analysis of molecular recognition: Steric electrostatic and hydrophobic complementarity. *Journal of Molecular Recognition*, **6**(4), pp.205–210.

Natesh, R., Schwager, S.L.U., Sturrock, E.D. and Acharya, K.R.,(2003). Crystal structure of the human angiotensin-converting enzyme–lisinopril complex. *Nature*, **421**(6922), pp.551–554.

Nielsen, H.M., Veerhuis, R., Holmqvist, B. and Janciauskiene, S.,(2009). Binding and uptake of Aβ1-42 by primary human astrocytes *in vitro*. *Glia*, **57**(9), pp.978–988.

Oda, T., Pasinetti, G.M., Osterburg, H.H., Anderson, C., Johnson, S.A. and Finch, C.E.,(1994). Purification and characterization of brain clusterin. *Biochemical and Biophysical Research Communications*, **204**(3), pp.1131–1136.

Oefner, C., D'Arcy, A., Hennig, M., Winkler, F.K. and Dale, G.E.,(2000). Structure of human

neutral endopeptidase (neprilysin) complexed with phosphoramidon. *Journal of Molecular Biology*, **296**(2), pp.341–349.

Oefner, C., Pierau, S., Schulz, H. and Dale, G.E.,(2007). Structural studies of a bifunctional inhibitor of neprilysin and DPP-IV. *Acta Crystallographica Section D*, **63**(9), pp.975–981.

Oefner, C., Roques, B.P., Fournie-Zaluski, M.-C. and Dale, G.E.,(2004). Structural analysis of neprilysin with various specific and potent inhibitors. *Acta Crystallographica Section D*, **60**(2), pp.392–396.

Oyama, F., Cairns, N.J., Shimada, H., Oyama, R., Titani, K. and Ihara, Y.,(1994). Down's syndrome: up-regulation of β-amyloid protein precursor and tau mRNAs and their defective coordination. *Journal of Neurochemistry*, **62**(3), pp.1062–1066.

Palomares, L.A., Estrada-Mondaca, S. and Ramírez, O.T.,(2004). Production of recombinant proteins: challenges and solutions. *Methods in Molecular Biology*, **267**, pp.15–51.

Pankow, K., Schwiebs, A., Becker, M., Siems, W.-E., Krause, G. and Walther, T.,(2009). Structural substrate conditions required for neutral endopeptidase-mediated natriuretic peptide degradation. *Journal of Molecular Biology*, **393**(2), pp.496–503.

Patterson, A.L.,(1935). A direct method for the determination of the components of interatomic distances in crystals. *Zeitschrift für Kristallographie - Crystalline Materials*, **90**(1–6), pp.517–542.

Pearson, H.A. and Peers, C.,(2006). Physiological roles for amyloid β peptides. *Journal of Physiology*, **575**(1), pp.5–10.

Pelmenschikov, V., Blomberg, M.R.A. and Siegbahn, P.E.M.,(2002). A theoretical study of the mechanism for peptide hydrolysis by thermolysin. *Journal of Biological Inorganic Chemistry*, **7**(3), pp.284–298.

Perrakis, A., Harkiolaki, M., Wilson, K.S. and Lamzin, V.S.,(2001). *ARP / wARP* and molecular replacement. *Acta Crystallographica Section D Biological Crystallography*, **57**(10), pp.1445–1450.

Perutz, M.F.,(1956). Isomorphous replacement and phase determination in non-centrosymmetric space groups. *Acta Crystallographica*, **9**(11), pp.867–873.

Plantinga, M.J., Korennykh, A. V, Piccirilli, J.A. and Correll, C.C.,(2008). Electrostatic interactions guide the active site face of a structure-specific ribonuclease to its RNA substrate. *Biochemistry*, **47**(34), pp.8912–8918.

Potter, L.R.,(2010). Guanylyl cyclases. In: R.A. Bradshaw, and E.A. Dennis. *Handbook of Cell Signaling*. Elsevier Inc. pp.1399–1407.

Potter, L.R.,(2011). Natriuretic peptide metabolism, clearance and degradation. *The FEBS Journal*, **278**(11), pp.1808-1817.

Powell, H.R., (2017). X-ray data processing. *Bioscience Reports*, **37**(5). Available from: https://portlandpress.com/bioscirep/article/37/5/BSR20170227/56808/X-ray-data-processing [Accessed Feburary 2020].

Prasher, V.P., Farrer, M.J., Kessling, A.M., Fisher, E.M.C., West, R.J., Barber, P.C. and Butler, A.C.,(1998). Molecular mapping of Alzheimer-type dementia in Down's syndrome. *Annals of Neurology*, **43**(3), pp.380–383.

Prince, M., Wimo, A., Guerchet, M., Gemma-Claire Ali, M., Wu, Y.-T., Prina, M., Yee Chan, K. and Xia, Z.,(2015). *World Alzheimer Report 2015: The global impact of dementia an analysis of prevalence, incidence, cost and trends*. London: Alzheimer's Disease International.

Qiu, W.Q., Walsh, D.M., Ye, Z., Vekrellis, K., Zhang, J., Podlisny, M.B., Rosner, M.R., Safavi, A., Hersh, L.B. and Selkoe, D.J.,(1998). Insulin-degrading enzyme regulates extracellular levels of amyloid beta-protein by degradation. *The Journal of Biological Chemistry*, **273**(49), pp.32730–32738.

Qiu, W.Q., Ye, Z., Kholodenko, D., Seubert, P. and Selkoe, D.J.,(1997). Degradation of amyloid beta-protein by a metalloprotease secreted by microglia and other neural and non-neural cells. *The Journal of Biological Chemistry*, **272**(10), pp.6641–6646.

Quay, T., Slaughter, C., Davis, T.P., Merrill, B.J. and Hersh, L.B.,(1994). Positional effects in the neprilysin (neutral endopeptidase) reaction. *Archives of Biochemistry and Biophysics*, **308**(1), pp.133–136.

R Core Team,(2019). *A language and environment for statistical computing* [computer

programme]. Available from: http://www.R-project.org/ [Accessed 2019].

Raffaella Tundo, G., Sbardella, D., Ciaccio, C., Bianculli, A., Orlandi, A., Desimio, M.G., Arcuri, G., Coletta, M. and Marini, S.,(2012). Insulin-degrading enzyme (IDE) a novel heat shock-like protein. *The Journal of Biological Chemistry*, **288**(4), pp.2281–2289.

Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V.,(1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, **7**(1), pp.95–99.

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A. and Finn, R.D.,(2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*, **46**(D1), pp.D624–D632.

Read, R.J.,(2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallographica. Section D, Biological Crystallography*, **57**(10), pp.1373–82.

Receveur-Brechot, V. and Durand, D.,(2012). How random are intrinsically disordered proteins? A small angle scattering perspective. *Current Protein & Peptide Science*, **13**(1), pp.55–75.

Renaud, J.-P., Chari, A., Ciferri, C., Liu, W., Rémigy, H.-W., Stark, H. and Wiesmann, C.,(2018). Cryo-EM in drug discovery: achievements, limitations and prospects. *Nature Reviews Drug Discovery*, **17**(7), pp.471–492.

Rogeberg, M., Furlund, C.B., Moe, M.K. and Fladby, T.,(2014). Identification of peptide products from enzymatic degradation of amyloid beta. *Biochimie*, **105**, pp.216–220.

Roher, A.E., Kasunic, T.C., Woods, A.S., Cotter, R.J., Ball, M.J. and Fridman, R.,(1994). Proteolysis of A beta peptide from Alzheimer disease brain by gelatinase A. *Biochemical and Biophysical Research Communications*, **205**(3), pp.1755–1761.

Rovelet-Lecrux, A., Hannequin, D., Raux, G., Meur, N. Le, Laquerrière, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M., Dubas, F., Frebourg, T. and Campion, D.,(2006). APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nature Genetics*, **38**(1), pp.24–26.

Rupp, B., (2010). *Biomolecular crystallography: principles, practice, and application to structural biology*. New York: Garland Science.

Sahli, S., Frank, B., Schweizer, W.B., Diederich, F., Blum-Kaelin, D., Aebi, J.D., Böhm, H., Oefner, C. and Dale, G.E.,(2005). Second-generation inhibitors for the metalloprotease neprilysin based on bicyclic heteroaromatic scaffolds: synthesis, biological activity, and X-ray crystal-structure analysis. *Helvetica Chimica Acta*, **88**(4), pp.731–750.

Saido, T. and Leissring, M.A.,(2012). Proteolytic degradation of amyloid β-protein. *Cold Spring Harbor Perspectives in Medicine*, **2**(6), Available from: http://perspectivesinmedicine.cshlp.org/content/2/6/a006379 [Accessed 2019].

Saito, T., Iwata, N., Tsubuki, S., Takaki, Y., Takano, J., Huang, S.M., Suemoto, T., Higuchi, M. and Saido, T.C.,(2005). Somatostatin regulates brain amyloid β peptide Aβ 42 through modulation of proteolytic degradation. *Nature Medicine*, **11**(4), pp.434–439.

Sammito, M., Meindl, K., de Ilarduya, I.M., Millán, C., Artola-Recolons, C., Hermoso, J.A. and Usón, I.,(2014). Structure solution with ARCIMBOLDO using fragments derived from distant homology models. *FEBS Journal*, **281**(18), pp.4029–4045.

Sato, B., Katagiri, Y.U., Iijima, K., Yamada, H., Ito, S., Kawasaki, N., Okita, H., Fujimoto, J. and Kiyokawa, N.,(2012). The human CD10 lacking an N-glycan at Asn628 is deficient in surface expression and neutral endopeptidase activity. *Biochimica et Biophysica Acta*, **1820**(11), pp.1715–1723.

Schechter, I. and Berger, A.,(1967). On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*, **27**(2), pp.157–162.

Schiering, N., D 'arcy, A., Villard, F., Ramage, P., Logel, C., Cumin, F., Ksander, G.M., Wiesmann, C., Karki, R.G. and Mogi, M.,(2016). Structure of neprilysin in complex with the active metabolite of sacubitril. *Scientific Reports*, **6**. Available from: https://www.nature.com/articles/srep27909 [Accessed 2019].

Schmidt, M., Rohou, A., Lasker, K., Yadav, J.K., Schiene-Fischer, C., Fändrich, M. and Grigorieff, N.,(2015). Peptide dimer structure in an Aβ(1-42) fibril visualized with cryo-EM. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(38), pp.11858–11863.

Schrödinger, L.,(2015). *The PyMOL Molecular Graphics System (Version 1.8)* [Computer programme]. Available from: https://pymol.org/2/ [Accessed 2019].

Schulz, H., Dale, G.E., Karimi-Nejad, Y. and Oefner, C.,(2009). Structure of human endothelin-converting enzyme I complexed with phosphoramidon. *Journal of Molecular Biology*, **385**(1), pp.178–187.

Schweizer, A., Valdenaire, O., Nelböck, P., Deuschle, U., Dumas Milne Edwards, J.B., Stumpf, J.G. and Löffler, B.M.,(1997). Human endothelin-converting enzyme (ECE-1): three isoforms with distinct subcellular localizations. *The Biochemical Journal*, **328**(3), pp.871–877.

Seed, A., Kuc, R.E., Maguire, J.J., Hillier, C., Johnston, F., Essers, H., de Voogd, H.J., McMurray, J. and Davenport, A.P.,(2012). The dual endothelin converting enzyme/neutral endopeptidase inhibitor SLV-306 (daglutril), inhibits systemic conversion of big endothelin-1 in humans. *Life Sciences*, **91**(13–14), pp.743–748.

Selkoe, D.J. and Hardy, J.,(2016). The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Molecular Medicine,* **8**(6), pp.595–608.

Sexton, T., Hitchcook, L.J., Rodgers, D.W., Bradley, L.H. and Hersh, L.B.,(2012). Active site mutations change the cleavage specificity of neprilysin. *PLoS ONE*, **7**(2), pp.1–10.

Shen, Y., Joachimiak, A., Rosner, M.R. and Tang, W.-J.,(2006). Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. *Nature*, **443**(7113), pp.870–874.

Shibata, M., Yamada, S., Kumar, S.R., Calero, M., Bading, J., Frangione, B., Holtzman, D.M., Miller, C.A., Strickland, D.K., Ghiso, J. and Zlokovic, B. V,(2000). Clearance of Alzheimer's amyloid-β(1-40) peptide from brain by LDL receptor-related protein-1 at the blood-brain barrier. *The Journal of Clinical Investigation*, **106**(12), pp.1489–1499.

Singh, S.P., Purohit, M.K., Aoyagi, C., Kitaoka, M. and Hayashi, K.,(2010). Effect of growth temperature, induction, and molecular chaperones on the solubilization of over-expressed cellobiose phosphorylase from cell V*ibrio gilvus* under *in vivo* conditions. *Biotechnology and Bioprocess Engineering*, **15**(2), pp.273–276.

Skidgel, R.A. and Erdös, E.G.,(1987). The broad substrate specificity of human angiotensin I converting enzyme. *Clinical and Experimental Hypertension. Part A: Theory and Practice*, **9**(2–3), pp.243–259.

Sleegers, K., Brouwers, N., Gijselinck, I., Theuns, J., Goossens, D., Wauters, J., Del-Favero, J., Cruts, M., Van Duijn, C.M. and Broeckhoven, C. Van,(2006). APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain*, **129**, pp.2977–2983.

Slunt, H.H., Thinakaran, G., Von Koch, C., Lo, A.C., Tanzi, R.E. and Sisodia, S.S.,(1994). Expression of a ubiquitous, cross-reactive homologue of the mouse β-amyloid precursor protein (APP). *Journal of Biological Chemistry*, **269**(4), pp.2637–2644.

Soba, P., Eggert, S., Wagner, K., Zentgraf, H., Siehl, K., Kreger, S., Löwer, A., Langer, A., Merdes, G., Paro, R., Masters, C.L., Müller, U., Kins, S. and Beyreuther, K.,(2005). Homo- and heterodimerization of APP family members promotes intercellular adhesion. *The EMBO Journal*, **24**(20), pp.3624–3634.

Spungin-Bialik, A., Ben-Meir, D., Fudim, E., Carmeli, S. and Blumberg, S.,(1996). Sensitive substrates for neprilysin (neutral endopeptidase) and thermolysin that are highly resistant to serine proteases. *FEBS Letters*, **380**(1–2), pp.79–82.

Stelzma, R.A., Schnitzlein, H.N. and Murllagh, F.R.,(1995). An English translation of 'ijber eine eigenartige Erlranliung der Hirnrinde &quot'. *Clinical Anatomy*, **8**, pp.429–443.

De Strooper, B.,(2014). Lessons from a failed γ-secretase Alzheimer trial. *Cell*, **159**(4), pp.721–726.

Sudoh, S., Frosch, M.P. and Wolf, B.A.,(2002). Differential effects of proteases involved in intracellular degradation of amyloid beta-protein between detergent-soluble and -insoluble pools in CHO-695 cells. *Biochemistry*, **41**(4), pp.1091–1099.

Takahashi, R.H., Milner, T.A., Li, F., Nam, E.E., Edgar, M.A., Yamaguchi, H., Beal, M.F., Xu, H., Greengard, P. and Gouras, G.K.,(2002). Intraneuronal Alzheimer Aβ42 accumulates in multivesicular bodies and is associated with synaptic pathology. *The American Journal of Pathology*, **161**(5), pp.2281–2289.

Takeda, S., Hashimoto, T., Roe, A.D., Hori, Y., Spires-Jones, T.L. and Hyman, B.T.,(2013). Brain interstitial oligomeric amyloid β increases with age and is resistant to clearance from brain in a mouse model of Alzheimer's disease. *The FASEB Journal*, **27**(8), pp.3239–3248.

Talafous, J., Marcinowski, K.J., Klopman, G. and Zagorski, M.G.,(1994). Solution Structure of Residues 1-28 of the Amyloid-Peptide. *Biochemistry*, **33**(25), pp.7788–7796.

Tamayev, R., Zhou, D. and D'adamio, L.,(2009). The interactome of the amyloid β precursor protein family members is shaped by phosphorylation of their intracellular domains. *Molecular Neurodegeneration*, **4**(28), pp.1–14.

Taylor, G.,(2003). The phase problem. *Acta Crystallographica Section D Biological Crystallography*, **59**(11), pp.1881–1890.

Teixeira, P.F., Masuyer, G., Pinho, C.M., Branca, R.M.M., Kmiec, B., Wallin, C., Wärmländer, S.K.T.S., Berntsson, R.P.-A., Ankarcrona, M., Gräslund, A., Lehtiö, J., Stenmark, P. and Glaser, E.,(2018). Mechanism of peptide binding and cleavage by the human mitochondrial peptidase neurolysin. *Journal of Molecular Biology*, **430**(3), pp.348–362.

Tschopp, J.F., Brust, P.F., Cregg, J.M., Stillman, C.A. and Gingeras, T.R.,(1987). Expression of the *lacZ* gene from two methanol-regulated promoters in *Pichia pastoris*. *Nucleic Acids Research*, **15**(9), pp.3859–3876.

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I.M., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C.A.K., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., Von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., Von Heijne, G., Nielsen, J. and Pontén, F.,(2015). Tissue-based map of the human proteome. *Science*, **347**(6220). Available from: https://science.sciencemag.org/content/347/6220/1260419 [Accessed 2019].

Usón, I. and Sheldrick, G.M.,(1999). Advances in direct methods for protein crystallography. *Current Opinion in Structural Biology*, **9**(5), pp.643–648.

Vagin, A. and Lebedev, A.,(2015). MoRDa , an automatic molecular replacement pipeline.

Acta Crystallographica Section A Foundations and Advances, **71**(1), p.S19.

Vagin, A. and Teplyakov, A.,(2010). Molecular replacement with *MOLREP*. *Acta Crystallographica Section D Biological Crystallography*, **66**(1), pp.22–25.

Valdenaire, O., Lepailleur-Enouf, D., Egidy, G., Thouard, A., Barret, A., Vranckx, R., Tougard, C. and Michel, J.-B.,(1999). A fourth isoform of endothelin-converting enzyme (ECE-1) is generated from an additional promoter. Molecular cloning and characterization. *European Journal of Biochemistry*, **264**(2), pp.341–349.

Valdenaire, O., Rohrbacher, E. and Mattei, M.G.,(1995). Organization of the gene encoding the human endothelin-converting enzyme (ECE-1). *The Journal of Biological Chemistry*, **270**(50), pp.29794–29798.

Vassar, R.,(1999). β-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science*, **286**(5440), pp.735–741.

Vassar, R.,(2001). The β-secretase, BACE: A prime drug target for Alzheimer's disease. *Journal of Molecular Neuroscience*, **17**(2), pp.157–170.

Vazeux, G., Wang, J., Corvol, P. and Llorens-Cortè, C.,(1996). Identification of glutamate residues essential for catalytic activity and zinc coordination in aminopeptidase A. *The Journal of Biological Chemistry*, **271**(15), pp.9069–9074.

Vekrellis, K., Ye, Z., Qiu, W.Q., Walsh, D., Hartley, D., Rie Chesneau, V., Rosner, M.R. and Selkoe, D.J.,(2000). Neurons regulate extracellular levels of amyloid-protein via proteolysis by insulin-degrading enzyme. *Journal of Neuroscience*, **20**(5), pp.1657–1665.

Vetrivel, K.S., Zhang, X., Meckler, X., Cheng, H., Lee, S., Gong, P., Lopes, K.O., Chen, Y., Iwata, N., Yin, K.-J., Lee, J.-M., Parent, A.T., Saido, T.C., Li, Y.-M., Sisodia, S.S. and Thinakaran, G.,(2008). Evidence that CD147 modulation of beta-amyloid (Abeta) levels is mediated by extracellular degradation of secreted Abeta. *The Journal of Biological Chemistry*, **283**(28), pp.19489–19498.

Viola, K.L. and Klein, W.L.,(2015). Amyloid β oligomers in Alzheimer's disease pathogenesis, treatment, and diagnosis. *Acta Neuropathologica*, **129**(2), pp.183–206.

Walsh, D.M., Klyubin, I., Fadeeva, J. V., Cullen, W.K., Anwyl, R., Wolfe, M.S., Rowan, M.J. and Selkoe, D.J.,(2002a). Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation *in vivo*. *Nature*, **416**(6880), pp.535–539.

Wälti, M.A., Ravotti, F., Arai, H., Glabe, C.G., Wall, J.S., Böckmann, A., Güntert, P., Meier, B.H. and Riek, R.,(2016). Atomic-resolution structure of a disease-relevant Aβ(1-42) amyloid fibril. *Proceedings of the National Academy of Sciences of the United States of America*, **113**(34), pp.4976-4984.

Wang, J., Brudvig, G.W., Batista, V.S. and Moore, P.B.,(2017). On the relationship between cumulative correlation coefficients and the quality of crystallographic data sets. *Protein Science*, **26**(12), pp.2410–2416.

Watanabe, Y., Nakajima, K., Shimamori, Y. and Fujimoto, Y.,(1997). Comparison of the hydrolysis of the three types of natriuretic peptides by human kidney neutral endopeptidase 24.11. *Biochemical and Molecular Medicine*, **61**(1), pp.47–51.

Waterman, D.G., Winter, G., Gildea, R.J., Parkhurst, J.M., Brewster, A.S., Sauter, N.K. and Evans, G.,(2016). Diffraction-geometry refinement in the DIALS framework. *Acta Crystallographica Section D Biological Crystallography*, **39**(4), pp.558–575.

Weidemann, A., Eggert, S., Reinhard, F.B.M., Vogel, M., Paliga, K., Baier, G., Masters, C.L., Beyreuther, K. and Evin, G.,(2002). A novel epsilon-cleavage within the transmembrane domain of the Alzheimer amyloid precursor protein demonstrates homology with Notch processing. *Biochemistry*, **41**(8), pp.2825–2835.

Weidemann, A., Kiinig, G., Bunke, D., Fischer, P., Salbaum, J.M., Masterst, C.L. and Beyreuther 'l, K.,(1989). Identification, biogenesis, and localization of precursors of Alzheimer's disease A4 amyloid protein. *Cell*, **57**(1), pp.115–126.

Weiss, M.S.,(2001). Global indicators of X-ray data quality. *Journal of Applied Crystallography*, **34**(2), pp.130–135.

Weiss, M.S. and Hilgenfeld, R.,(1997). *On the use of the merging R factor as a quality indicator for X-ray data*. *Journal of Applied Crystallography*, **30**(2), pp.203–205.

Weller, R.O., Massey, A., Kuo, Y.-M. and Roher, A.E.,(2000). Cerebral myloid angiopathy:

accumulation of Abeta in interstitial fluid drainage pathways in Alzheimer's disease. *Annals of the New York Academy of Sciences*, **903**(1), pp.110–117.

Wiedemann, H.,(2015). Overview of Synchrotron Radiation. In: H, Wiedemann. Particle Accelerator Physics. Springer. pp.815–855.

Williams, T.L., Serpell, L.C. and Urbanc, B.,(2016). Stabilization of native amyloid β-protein oligomers by copper and hydrogen peroxide induced cross-linking of unmodified proteins (CHICUP). *Biochimica et Biophysica Acta – Proteins and Proteomics*, **1864**(3), pp.249–259.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S. and IUCr,(2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography*, **67**(4), pp.235–242.

Winter, G.,(2010). *xia2* : an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, **43**(1), pp.186–190.

Winter, G. and McAuley, K.E.,(2011). Automated data collection for macromolecular crystallography. *Methods*, **55**(1), pp.81–93.

Wlodawer, A., Minor, W., Dauter, Z. and Jaskolski, M.,(2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS Journal*, **275**(1), pp.1–21.

Wood, J.G., Mirra, S.S., Pollock, N.J. and Binder, L.I.,(1986). Neurofibrillary tangles of Alzheimer disease share antigenic determinants with the axonal microtubule-associated protein tau (τ). *Proceedings of the National Academy of Sciences of the United States of America*, **83**(11), pp.4040–4043.

Woods Group,(2019). *GLYCAM Web* [computer programme]. Available from: http://glycam.org [Accessed 2016].

Wyss-Coray, T., Loike, J.D., Brionne, T.C., Lu, E., Anankov, R., Yan, F., Silverstein, S.C. and Husemann, J.,(2003). Adult mouse astrocytes degrade amyloid-β *in vitro* and *in situ*.

Nature Medicine, **9**(4), pp.453–457.

Xiao, Y., Ma, B., McElheny, D., Parthasarathy, S., Long, F., Hoshi, M., Nussinov, R. and Ishii, Y.,(2015). Aβ(1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nature Structural & Molecular Biology*, **22**(6), pp.499–505.

Xu, D., Emoto, N., Giaid, A., Slaughter, C., Kaw, S., DeWit, D. and Yanagisawa, M.,(1994). ECE-1: A membrane-bound metalloprotease that catalyzes the proteolytic activation of big endothelin-1. *Cell*, **78**(3), pp.473–485.

Xu, J., Carretero, O.A., Liu, Y.-H., Yang, F., Shesely, E.G., Oja-Tebbe, N. and Yang, X.-P.,(2004). Dual inhibition of ACE and NEP provides greater cardioprotection in mice with heart failure. *Journal of Cardiac Failure*, **10**(1), pp.83–89.

Yang, M. and Teplow, D.B.,(2008). Amyloid β-[rotein monomer folding: free-energy surfaces reveal alloform-specific differences. *Journal of Molecular Biology*, **384**(2), pp.450–464.

Young-Pearse, T.L., Bai, J., Chang, R., Zheng, J.B., Loturco, J.J. and Selkoe, D.J.,(2007). A critical function for-amyloid precursor protein in neuronal migration revealed by in utero RNA interference. *Journal of Neuroscience*, **27**(52), pp.14459–14469.

Yu, H.,(1999). Extending the size limit of protein nuclear magnetic resonance. *Proceedings of the National Academy of Sciences of the United States of America*, **96**(2), pp.332–334.

Zhang, Z., Witham, S. and Alexov, E.,(2011). On the role of electrostatics in protein-protein interactions. *Physical Biology*, **8**(3). Available from: https://iopscience.iop.org/article/10.1088/1478-3975/8/3/035001 [Accessed 2019].

Zheng, L., Calvo-Garrido, J., Hallbeck, M., Hultenby, K., Marcusson, J., Cedazo-Minguez, A. and Terman, A.,(2013). Intracellular localization of amyloid-peptide in SH-SY5Y neuroblastoma cells. *Journal of Alzheimer's Disease*, **37**(4), pp.713–733.

Zhou, H.-X. and Pang, X.,(2018). Electrostatic interactions in protein structure, folding, binding, and condensation. *Chemical Reviews*, **118**(4), pp.1691–1741.

Zois, N.E., Bartels, E.D., Hunter, I., Kousholt, B.S., Olsen, L.H. and Goetze, J.P.,(2014). Natriuretic peptides in cardiometabolic regulation and disease. *Nature Reviews*

Cardiology, **11**(7), pp.403–412.